

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: December 10, 2003, 18:29:51 ; Search time 20.9215 Seconds
(without alignments)
1776.146 Million cell updates/sec

Title: US-10-080-113-4
Perfect score: 711
Sequence: 1 MNISVNPYLMVAVVVFVLL.....AQMVFEDLNKRVANGLS 144

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL 23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaea:*

ALIGNMENTS

RESULT 1

O25757 PRELIMINARY; PRT; 144 AA.

AC O25757; DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE ATP synthase F0, subunit B (ATPF).
GN HPI137.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
RX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Hickley A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
pylori";
RL Nature 388:539-547(1997).
DR EMBL; AE000619; AAD08179.1; -.
DR TIGR; HP1137; -.
DR InterPro; IPR002146; ATPsynt B/B'sub.
DR Pfam; PF00430; ATP-synt B; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 144 AA; 16497 MW; D351321DFC5M4209 CRC64;

Query Match 100.0%; Score 711; DB 16; Length 144;
Best Local Similarity 100.0%; Pred. No. 5.2e-43;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

SUMMARIES				Description	
Result No.	Score	Query Match	Length DB ID		
1	711	100.0	144 16 O25757	O25757 helicobacte	
2	685	96.3	144 16 Q92K76	Q92K76 helicobacte	
3	169	23.8	141 16 Q9FJ24	Q9FJ24 campylobact	
4	146	20.5	219 10 Q42139	Q42139 arabidopsis	
5	141	19.8	138 16 Q8LDP6	Q8LDP6 synechococc	
6	138	19.4	219 10 Q94B39	Q94B39 arabidopsis	
7	132	18.6	1179 17 Q59462	Q59462 pyrococcus	
8	127	17.9	163 2 Q8KRV2	Q8KRV2 illyobacter	
9	127	17.9	164 16 Q8XIU9	Q8XIU9 thermotoga	
10	125	17.6	156 16 Q8XGD7	Q8XGD7 salmonella	
11	121.5	17.1	156 16 Q8XU72	Q8XU72 ralsstonia s	
12	120.5	16.9	156 16 Q9HT16	Q9HT16 pseudomonas	
13	120	16.8	156 16 Q8Z982	Q8Z982 versinia pe	
14	119.5	16.8	144 16 Q67525	Q67525 aquifex aeo	
15	119.5	16.8	156 16 Q9PE81	Q9PE81 xyliella fas	
16	119	16.7	168 16 Q8RC19	Q8RC19 thermoanaer	

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```

QY 1 MNISVNPYLMAVVFVLLWMNVVYRPLLAFMDNRQAEIKDSLAKIKTNDNAQSVEI 60
DB 1 MNISVNPYLMAVVFVLLWMNVVYRPLLAFMDNRQAEIKDSLAKIKTNDNAQSVEI 60
QY 61 GHOIEALLKEAAEKREIIIAEIAIKATESYDAVIKOKENELNQEFPAFAKOLQNEKQALK 120
DB 61 GHOIEALLKEAAEKREIIIAEIAIKATESYDAVIKOKENELNQEFPAFAKOLQNEKQALK 120
QY 121 EQLQAMQPVFEDELNKRVAAGLGS 144
DB 121 EQLQAMQPVFEDELNKRVAAGLGS 144

RESULT 2
Q9ZK76
ID Q9ZK76 PRELIMINARY; PRT; 144 AA.
AC Q9ZK76
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE ATP synthase B'.
GN ATPX OR JHP1065.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
DR EMBL, AE001533, RAD06630.1; -.
DR InterPro; IPR002146; ATPsynth_B/B'sub.
DR Pfam; PF00430; ATP-synt_B; 1.
KW Complete proteome.
SQ SEQUENCE 144 AA; 16621 MW; D584D48E526EF09D CRC64;

Query Match 96.3%; Score 685; DB 16; Length 144;
Best Local Similarity 95.8%; Pred No. 3,5e-41;
Matches 138; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNISVNPYLMAVVFVLLWMNVVYRPLLAFMDNRQAEIKDSLAKIKTNDNAQSVEI 60
DB 1 MNISVNPYLMAVVFVLLWMNVVYRPLLAFMDNRQAEIKDSLAKIKTNDNAQSVEI 60
QY 61 GHOIEALLKEAAEKREIIIAEIAIKATESYDAVIKOKENELNQEFPAFAKOLQNEKQALK 120
DB 61 GHOIEALLKEAAEKREIIIAEIAIKATESYDAVIKOKENELNQEFPAFAKOLQNEKQALK 120
QY 121 EQLQAMQPVFEDELNKRVAAGLGS 144
DB 121 EQLQAMQPVFEDELNKRVAAGLGS 144

RESULT 3
Q9PJ24
ID Q9PJ24 PRELIMINARY; PRT; 141 AA.
AC Q9PJ24
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE ATP synthase F0 sector B' subunit (EC 3.6.1.34).
GN ATPF OR CJO102.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.

```

```

OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NCCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Kren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham K., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,
RA Jagers K., Karyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139074; CAB72586.1; -.
DR InterPro; IPR002146; ATPsynth_B/B'sub.
DR Pfam; PF00430; ATP-synt_B; 1.
KW Complete proteome.
SQ SEQUENCE 141 AA; 16352 MW; 664A478F8DDEDF5 CRC64;

Query Match 23.8%; Score 169; DB 16; Length 141;
Best Local Similarity 25.4%; Pred No. 7,7e-05;
Matches 34; Conservative 37; Mismatches 63; Indels 0; Gaps 0;

QY 5 VNPYLMAVVFVLLWMNVVYRPLLAFMDNRQAEIKDSLAKIKTNDNAQSVEIGHQI 64
DB 5 MHPSIMLATWATFLAMIVILNSMLYKPLLKFMDERNDSIKNDENKVSQEVLGNDL 64
QY 65 EALLKEAAEKREIIIAEIAIKATESYDAVIKOKENELNQEFPAFAKOLQNEKQALK 124
DB 65 EAHINTRREIIOKIOSAIAAAEAEQILRSKELEKQASFYADLAVQKLEQHLN 124
QY 125 AQMPVFEDELNKR 138
DB 125 IHLSELKQALONNI 138

RESULT 4
Q42139
ID Q42139 PRELIMINARY; PRT; 219 AA.
AC Q42139
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE H+-transporting ATP synthase CHAIN9 - like protein
DE (ATG32260/F10M6_100).
GN ATPG OR F10M6_100 OR ATG32260.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Weichselgartner M., Fartmann B., Granderath K., Dauner D.,
RA Herzl A., Neumann S., Hoheisel J., Mewes H.W., Mayer K., Schueller C.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 155-219 FROM N.A.
RX STRAIN=cv. Columbia; TISSUE=Seedling;
RA Desprez T., Amselem J., Chispello H., Caboche M., Hofte H.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Columbia;
RA Legen J., Misera S., Herrmann R.G., Altschmid L.;
RT "Sequences and map position of 31 Arabidopsis thaliana cDNAs encoding
RT organellar polypeptides.";
RN [4]
RP Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Columbia;
RX MEDLINE=95148729; PubMed=7846151;
RA Newman T., de Bruijn F.J., Green P., Keegstra K., Kende H.,

```


[illegible]

```

DR Pfam: PF00430; ATP-synt_E; 1.
SQ SEQUENCE 219 AA; 23948 MW; ABB29BB03E84B5DE CRC64;

Query Match 19.4%; Score 138; DS 10; Length 219;
Best Local Similarity 24.8%; Pred. No. 0.019;
Matches 33; Conservative 36; Mismatches 56; Indels 8; Gaps 2;

QY 10 MAVVTVVLLINMNVVTRPLAFMNDNRQAIKDSLAKIKTDNAQSVETGHQIEALLK 69
DB 88 LPIIVVEFLFMFALDKVYISQLGNFMDQDASIKELKASVKDTSFEVKELDEQAAAYVR 147
QY 70 EAAEKRRBIIAEATKATSYDAVIKOK----ENELNQEFAPAKOLNEKALKEQIQQA 125
DB 148 AA-----RAIIAALNKKKKTQVEVEEKLAEGKKVVEELKEALSLSEKQETIKALDS 203
QY 126 QMPVFEDELNKRVR 138
DB 204 QIAALSEDIVKRV 216

RESULT 7
CS9462 PRELIMINARY; PRT; 1179 AA.
ID ID 059462; AC
OS 059462; AC
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-WAR-2003 (TReMBLrel. 23, Last annotation update)
DE 1179AA long hypochemical chromosome assembly protein.
PH1798.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1] _TaxID=53953;
RP RC SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=983441137; PubMed=3679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine K., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL: AF000007; BAA30917.1; -
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR005289; GTP-binding_dom.
DR InterPro; IPR003405; SMC C.
DR InterPro; IPR003395; SMC N.
DR InterPro; IPR002017; Spectrin.
DR Pfam: PF02483; SMC C; 1.
DR Pfam: PF02463; SMC N; 1.
DR TIGRFAMs: TIGR00650; MG442; 1.
DR Complete proteome.
SQ SEQUENCE 1179 AA; 135656 MW; 24A61BE5F3864493 CRC64;

Query Match 18.6%; Score 132; DB 17; Length 1179;
Best Local Similarity 31.5%; Pred. No. 0.29;
Matches 41; Conservative 25; Mismatches 46; Indels 18; Gaps 4;

QY 25 NVWVTRPLAFMNDNRQAIKDSLAKIKTDNAQSVETGHQIEALLKEAAEKRRB--- 77
DB 313 NTEIARRELDESQRLAKVDELKKNVSEIKSGKAIIRWGRREALIKQISEKEEENH 372
QY 78 -----ITAEAIQKATSYDAVIKOKEN--ELANCFEAPAKOLNEKALKEQIQQA 126
DB 373 LVVRIGEIDTRFVAREEFDVSVKLENAERLMEYGEAEIKRLDAEKEKLRSIAVLKAK 432
QY 127 MFVFEDELNK 136
DB 433 LFGRIDEILK 442

```

```

DR  f3am; PF00430; ATP-synt_B; 1.
DR  TIGRFAMs; TIGR01144; ATP_synt_b; 1.
KW  Complete proteome.
SQ  SEQUENCE 164 AA; 13295 MW; 7B0F14F285A88CE0 CRC64;

Query Match
Best Local Similarity 17.9%; Score 127; DB 16; Length 164;
Matches 32; Conservative 44; Mismatches 58; Indels 14; Gaps 4;

QY  3 ISVNPYLVAVFVVFVLLWMNVVYRPLLAFLPMNDRQAEIKDSLAKIKTDNAQSVEIGH 62
Db  4 LEIN-WTSAAMLVFLVWVFLNKLTPFIENAEKRRKKVBDLSAQBLKEAEKMS 62
QY  63 QIEALLKEAAEKREII-----AEAI-----QKATESYDAVIKOKENLNQFEAFKQ 111
Db  63 EAERFLSEARQADEIVESARKEAIVEAREKAKKEAQNIVESAKTOIEVYKKALEQ 122
QY  112 LQNEKQALKQEQQAQM--PVFEDELNKR 137
Db  123 VQRAAELSVILATKLQKQVQDERARR 150

RESULT 10
Q8XGD7 PRELIMINARY; PRT; 156 AA.
AC Q8XGD7;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE Membrane-bound ATP synthase, F0 sector, subunit b (EC 3.6.3.14) (ATP synthase subunit B).
DE ATPF OR STM3869 OR STV3909.
GN Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OC NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Ngwenlik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2."
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churche C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatia M., Baker S., Basham D., Brooks K., Chillingworth T., Connerton P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Feltwell T., Hanlin N., Haque A., Hien T.T., Holroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
DR EMBL; AE008880; AAL22727.1; -
DR EMBL; AL627280; CAD03126.1; -
DR InterPro; IPR005864; ATPaseB.
DR Pfam; PF00430; ATP-synt_B; 1.
DR TIGRFAMs; TIGR01144; ATP_synt_b; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 156 AA; 17365 MW; 46C5209A95BE160A CRC64;

Query Match
Best Local Similarity 17.1%; Score 121.5; DB 16; Length 156;
Matches 37; Conservative 35; Mismatches 52; Indels 21; Gaps 4;

QY  3 ISVNPYLVAVFVVFVLLWMNVVYRPLLAFLPMNDRQAEIKDSLAKIKTDNAQSVEIGH 62
Db  1 MNLNATLV AQM-VVFILWVWVAKFIWPLVKALDERAKKIADGLAARDKGALELANK 59
QY  63 QIEALLKEA-----AKREIIAEATOKATESYDA-VIKOKENLNQFEAFKQ 111
Db  60 RVEQALTEARTEGAQRIADAERAKQMTADEIKQNAQAARIVAAQAKAEQAVRVRES 119
QY  112 LQNE-----KQALKEQLQAQM 127
Db  120 LRDQAVLVAKGAEQILKREVNQAV 144

RESULT 12
Q9HT16 PRELIMINARY; PRT; 156 AA.
ID Q9HT16
AC Q9HT16;
DT 01-MAR-2001 (TREMBlrel. 16, Created)

```

```

DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE ATP synthase B chain.
GN ATPF OR PA558.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Gollry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laibig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL: AE004967; AAG08943.1; -.
DR HSSP: P00859; 1B3U.
DR InterPro: IPR005864; ATPaseB.
DR Pfam: PF00430; ATP-synt_B; 1.
DR TIGRFAMs: TIGR01144; ATP_synt_b; 1.
KW Complete proteome.
SQ SEQUENCE 156 AA; 16956 MW; F2B9EDF6A667761B CRC64;

Query Match 16.9%; Score 120.5; DB 16; Length 156;
Best Local Similarity 28.1%; Pred. No. 0.22;
Matches 38; Conservative 27; Mismatches 49; Indels 21; Gaps 5;

QY 1 MNISVNPYLMAVVFVVFVLLWMNVYRPLLAFLMNDNRQAEIKDSIAKIKTDNAQSVET 60
DB 1 MNINATLIGQSVAFVIFVFCMK---FVWPVIALQERQKIADGL-DAANRAARDLEL 56

QY 61 HQICBALKEAKREIRIIAIOKATESYDAVIKQKENELNQEFPAKOLQNEKQALK 120
DB 57 AHE-----KAGQQLREAKQAQAE-----IVEQAKRANQIVDEARDQARTGERBLK 102

QY 121 EQLQKQMPVFEDLN 135
DB 103 AQAQAEI---EQLN 114

RESULT 13
Q8Z9S2 PRELIMINARY; PRT; 156 AA.
AC Q8Z9S2;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE ATP synthase subunit B protein (EC 3.6.1.34) (Membrane-bound ATP
DE synthase, F0 sector, subunit b).
GN ATPF OR YP04125 OR Y4139.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RA "Genome sequence of Yersinia pestis, the causative agent of plague.";

```

```

RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Petherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).
DR EMBL: AJ414160; CAC93574.1; -.
DR EMBL: AB014015; AAM87681.1; -.
DR InterPro: IPR005864; ATPaseB.
DR InterPro: IPR002146; ATP_synt_B/B'sub.
DR Pfam: PF00430; ATP-synt_B; 1.
DR TIGRFAMs: TIGR01144; ATP_synt_b; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 156 AA; 17258 MW; 6ECC0D738E2CC516 CRC64;

Query Match 16.9%; Score 120; DB 16; Length 156;
Best Local Similarity 24.5%; Pred. No. 0.24;
Matches 37; Conservative 36; Mismatches 42; Indels 36; Gaps 5;

QY 1 MNISVNPYLMAVVFVVFVLLWMNVYRPLLAFLMNDNRQAEIKDSL-----AKIKTDNA 55
DB 1 MNLNATLIGQIAFVLFVIFCMK---YVWPPIAAIEKRQOEIADGLSSAEAKKDLDLA 57

QY 56 QS-----VEIGHQIEALLKEAEKREIRIIAIOKATESYDAVIKQKENELNQEFAPA 109
DB 58 QANATDQLKAKAEQVIFQASKRAQLDQAKAEQERNKIVAQQAQAEIDAE----- 112

QY 110 KOLQNEKQALKEQLQQAQMPVFEDLNKRYAM 140
DB 113 -----RKRA-----REELRKQVAM 126

RESULT 14.
O67525 PRELIMINARY; PRT; 144 AA.
AC O67525;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE ATP synthase F0 subunit B.
GN ATPF1 OR AQ 1586.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Auja M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RA "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
DR EMBL: AE000746; AAC07477.1; -.
DR InterPro: IPR002146; ATP_synt_B/B'sub.
DR Pfam: PF00430; ATP-synt_B; 1.
KW Complete proteome.
SQ SEQUENCE 144 AA; 16719 MW; 7D5377479D5B0D2B CRC64;

Query Match 16.8%; Score 119.5; DB 16; Length 144;
Best Local Similarity 21.1%; Pred. No. 0.24;
Matches 30; Conservative 48; Mismatches 57; Indels 7; Gaps 3;

QY 1 MNISVNP-----YLMVAVFVFVLLWMNVYRPLLAFLMNDNRQAEIKDSIAKIKTDNAQ 56
DB 1 MGIWNPNTILVQLFIFVIFLMII--TNIVY-KYTVAVISREELIKKNVSEAQKREE 57

```

[illegible]

```

Query Match      16.8%; Score 119.5; DB 16; Length 156;
Best Local Similarity 26.9%; Pred. No. 0.26;
Matches 33; Conservative 23; Mismatches 51; Indels 13; Gaps 2;

Qy      10 MAVFVV-----PVLILMAMVWVYPLAFMDNQAEIKDSLAKIKTIDNAQSVIECHQ 63
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1 MDITFTFASLAPALINIVATKIWPLIKVIERQOKIAGLGAADLQKELAQAE 60
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      64 IEALLKAAKRRRIIAETAIQATSDYAVIKQENTLNQ-----EFEAFKQLONE 115.
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db      61 IKKTLNAREKANEIIEQAHARAHQIIEAAKGAETETNRQONLAQVEIEAAAKRARE 119

Search completed: December 10, 2003, 18:36:47
Job time : 24.9215 secs

```

7

45%

40%

35%

30%

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:28:11 ; Search time 44.9764 Seconds
(without alignments)
878.747 Million cell updates/sec

Title: US-10-080-113-1

Perfect score: 1254

Sequence: 1 MKGSLAIVLGSLLASGAFY.....LSKKHMEHYEKVSSEMKK 249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq.19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1254	100.0	249	19 AAW98332	H. pylori GHPO 161
2	1199	95.6	249	18 AAW55445	H. pylori ORF 04ae
3	1199	95.6	249	20 AAV17212	H. pylori outer me
4	1199	95.6	261	18 AAW55332	H. pylori ORF hp3e
5	642	51.2	132	23 ABUS1422	Helicobacter pylori
6	300	23.9	282	19 AAV10951	H. pylori ORF 04ee
7	300	23.9	282	20 AAV17214	H. pylori outer me
8	300	23.9	282	23 AAE15838	Helicobacter pylori
9	300	23.9	285	18 AAW20938	H. pylori secreted

10	300	23.9	285	23 AAE15839	Helicobacter pylori
11	297.5	23.7	254	21 AAE21214	Helicobacter pylori
12	297.5	23.7	254	23 AAE15841	Helicobacter pylori
13	297.5	23.7	396	21 AAE21215	Cyt/HPA44/HF fusio
14	204	16.3	337	21 AAO20108	Protein #1 encoded
15	204	16.3	461	19 AAW74466	Adhesin/V.cholerae
16	204	16.3	461	20 AAW32094	Adhesin/CTXA2B chi
17	201.5	16.1	260	18 AAW07450	Helicobacter pylori
18	201.5	16.1	260	19 AAW60157	Helicobacter pylori
19	201.5	16.1	260	19 AAY49540	Helicobacter pylori
20	201.5	16.1	260	21 AAY71926	Helicobacter pylori
21	201.5	16.1	260	23 AAE15834	Helicobacter pylori
22	199.5	15.9	260	18 AAW07449	Helicobacter pylori
23	199.5	15.9	260	19 AAW60156	Helicobacter pylori
24	199.5	15.9	260	20 AAY49539	Helicobacter pylori
25	199.5	15.9	260	21 AAY71925	Helicobacter pylori
26	199.5	15.9	260	23 AAE15833	Helicobacter pylori
27	194.5	15.5	260	19 AAY11000	H. pylori ORF 11ap
28	194.5	15.5	260	20 AAY71923	H. pylori outer me
29	194.5	15.5	260	21 AAY71927	Helicobacter pylori
30	194.5	15.5	260	23 AAE15835	Helicobacter pylori
31	190.5	15.2	258	18 AAW20973	H. pylori derived
32	190.5	15.2	258	21 AAY71928	Helicobacter pylori
33	190.5	15.2	258	23 AAE15836	Helicobacter pylori
34	171	13.6	137	23 ABUS1496	Helicobacter pylori
35	155	12.4	166	23 AAE15840	Helicobacter pylori
36	143.5	11.4	147	23 ABUS2240	Helicobacter pylori
37	134	10.7	147	18 AAW20468	H. pylori secreted
38	132	10.5	207	19 AAW98223	H. pylori GHPO 603
39	131	10.4	205	13 AAE21614	Sequence of plasma
40	129	10.3	100	23 ABUS1864	Helicobacter pylori
41	129	10.3	207	18 AAW55459	H. pylori ORF 04gp
42	129	10.3	218	18 AAW55202	H. pylori ORF 01cp
43	116.5	9.3	6815	22 ABB66811	Drosophila melanog
44	113.5	9.1	535	22 AAB92542	Human protein sequ
45	113.5	9.1	1373	24 ABUS11772	Human MDDT polypep

ALIGNMENTS

RESULT 1

AAW98332 ID AAW98332 standard; Protein; 249 AA.

AC AAW98332;

XX 31-MAR-1999 (first entry)

DT H. pylori GHPO 1615 protein.

DE GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease.

XX Helicobacter pylori.

XX WOS843478-A1.

PD 08-OCT-1998.

PF 01-APR-1998; 98WO-US06371.

PR 29-JUL-1997; 97US-0902615.

PR 01-APR-1997; 97US-0833457.

PR 24-JUN-1997; 97US-0881227.

XX (HUMA-) HUMAN GENOME SCI INC.
XX (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

FI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;

XX WPI; 1998-542293/46.

DR N-PSDB; AAX14051.

XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
PS Claim 8; Page 455-456; 2054pp; English.
XX
XX This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 249 AA;
Query Match 100.0%; Score 1254; DB 19; Length 249;
Best Local Similarity 100.0%; Pred. No. 4.1e-107;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MKKGLAIVLGSLLASGAFYALADGMPAKQOHNNTGESVELHFHYPIKQKQEPKNSHLV 60
Db 1 MKKGLAIVLGSLLASGAFYALADGMPAKQOHNNTGESVELHFHYPIKQKQEPKNSHLV 60
Qy 61 VLIEPKIEINKVIPESYOKEFEKSLFLQLSSFLERKGYVSQPKDASEIPQDIKEKALLV 120
Db 61 VLIEPKIEINKVIPESYOKEFEKSLFLQLSSFLERKGYVSQPKDASEIPQDIKEKALLV 120
Qy 121 LRMDGNVAILEDIVESDALSEEKVIDMSSGYLNLFNVEPKSEDIHSGIDVSKIKAVI 180
Db 121 LRMDGNVAILEDIVESDALSEEKVIDMSSGYLNLFNVEPKSEDIHSGIDVSKIKAVI 180
Qy 181 ERVELRRTNSSGGFVPKTFVHRKETDHDQAIRKIMNQAYHKVMVHITKELSKKHEHYEK 240
Db 181 ERVELRRTNSSGGFVPKTFVHRKETDHDQAIRKIMNQAYHKVMVHITKELSKKHEHYEK 240
Qy 241 VSSEMKRK 249
Db 241 VSSEMKRK 249
RESULT 2
AAW55445 standard; Protein; 249 AA.
ID AAW55445
XX AC AAW55445;
XX
XX 24-JUN-1998 (first entry)
XX
DE H. pylori ORF 01ael1010_40688_c2_38 cell envelope OMP.
XX
XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
XX
OS Helicobacter pylori.
XX
XX WO9737044-A1.
PN
PD 09-OCT-1997.
XX
XX 27-MAR-1997; 97WO-US05223.
PF
XX 06-DEC-1996; 96US-0761318.
PR 29-MAR-1996; 96US-0625811.
PR 02-APR-1996; 96US-0758731.
PR 25-OCT-1996; 96US-0736905.
PR 28-OCT-1996; 96US-0738859.
XX
XX (ASTR) ASTRA AB.
PA
XX Alm RA, Smith D;
PI

XX
DR WPI; 1997-503122/46.
DR N-PSDB; AAV24854.
XX
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX
PS Claims 14,80; Page 653; 1145pp; English.
XX
CC This sequence is a H. pylori cell envelope outer membrane protein (OMP).
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
XX
SQ Sequence 249 AA;
Query Match 95.6%; Score 1199; DB 18; Length 249;
Best Local Similarity 94.8%; Pred. No. 4.7e-102;
Matches 236; Conservative 6; Mismatches 7; Indels 0; Gaps 0;
Qy 1 MKKGLAIVLGSLLASGAFYALADGMPAKQOHNNTGESVELHFHYPIKQKQEPKNSHLV 60
Db 1 MKKGLAIVLGSLLASGAFYALADGMPAKQOHNNTGESVELHFHYPIKQKQEPKNSHLV 60
Qy 61 VLIEPKIEINKVIPESYOKEFEKSLFLQLSSFLERKGYVSQPKDASEIPQDIKEKALLV 120
Db 61 VLIDPKIEANKVIPENYOKEFEKSLFLQLSNFLERKGYVSQPKDASEIPQDIKEKALLV 120
Qy 121 LRMDGNVAILEDIVESDALSEEKVIDMSSGYLNLFNVEPKSEDIHSGIDVSKIKAVI 180
Db 121 LRMDGNVAILEDIVESDALSEEKVIDMSSGYLNLFNVEPKSEDIHSGIDVSKIKAVI 180
Qy 181 ERVELRRTNSSGGFVPKTFVHRKETDHDQAIRKIMNQAYHKVMVHITKELSKKHEHYEK 240
Db 181 ERVELRRTNSSGGFVPKTFVHRKETDHDQAIRKIMNQAYHKVMVHITKELSKKHEHYEK 240
Qy 241 VSSEMKRK 249
Db 241 VSSEMKRK 249
RESULT 3
AAV17212 standard; Protein; 249 AA.
ID AAV17212
XX AC AAV17212;
XX
XX 03-AUG-1999 (first entry)
XX
XX H. pylori outer membrane polypeptide.
DE Outer membrane polypeptide; OMP; vaccine; H. pylori infection; humoral;
XX cellular immune response.
XX
XX Helicobacter pylori.
XX
XX

PN WO921959-A2.
XX
PD 06-MAY-1999.
XX
PF 28-OCT-1998; 98WO-US22883.
XX
PR 17-DEC-1997; 97US-0993001.
PR 28-OCT-1997; 97US-0959131.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
FA
XX Alm RA, Ellis RW, Guild BC, Noonan BM, Smith D;
PI WPI; 1999-326698/27.
XX N-PSDB; AAX75831.
DR
XX Cellular vaccine against Helicobacter pylori
PT
XX
XX Claim 7; Page 304-305; 352pp; English.
PS
XX The invention relates to a vaccine for preventing or treating infections
CC by Helicobacter pylori. The vaccine contains at least one isolated
CC H. pylori polypeptide, or its fragments, in a carrier, where the
CC carrier is a Salmonella, Vibrio cholerae or Shigella vector containing a
CC nucleic acid encoding the H. pylori polypeptide. The vaccines induce
CC humoral and cellular immune responses. The vaccines are used to treat or
CC prevent infections by H. pylori. Sequences AAX75779 to AAX75837 represent
CC nucleic acid sequences encoding H. pylori outer membrane polypeptides
CC (OMPs) AAX17160 to AAX17218.
CC
XX
XX Sequence 249 AA;
SQ
Query Match 95.6%; Score 1199; DB 20; Length 249;
Best Local Similarity 94.8%; Pred. No. 4.7e-102;
Matches 236; Conservative 6; Mismatches 7; Indels 0; Gaps 0;
QY 1 MKGSLAIVLGSLLASGAFYTLADGMPAKOQHNTGESVELHFHYPIKQKQEPKNSHLV 60
DB 1 MKGSLAIVLGSLLASGTFYTLADGMPKQOHNNGESVELHFHYPIKQKQEPKNNHLV 60
QY 61 VLIEPKIEINKVIPESYQKEFEKSLFLQSLFLERKGYVSQFQKASEIPQDIKEKALLV 120
DB 61 VLIDPKIEANKVIPENYQKEFEKSLFLQSLNFLERKGYVSQFQKASEIPQDIKEKALLV 120
QY 121 LRMDGNVAILEDIVEESDALSEKVIDMSSGYLNLNFVEPKSEDIHHSFGIDVSKIKAVI 180
DB 121 LRMDGNVAILEDIVEESDALSEKVIDMSSGYLNLNFVEPKSEDIHHSFGIDVSKIKAVI 180
QY 181 ERVELARTNSGGFVPKTFVHRKETDHDQAIKIMNQAYHKVMVHITKELSKKHMEHYEK 240
DB 181 ERVELARTNSGGFVPKTFVHRKETDHDRAIKIMNQAYHKVMVHITKELSKKHMEHYEK 240
QY 241 VSSEMKKKR 249
DB 241 VSSEMKKKR 249
RESULT 4
AAW55332
ID AAW55332 standard; Protein; 261 AA.
XX
XX AAW55332;
XX
XX 15-JUN-1998 (first entry)
DT
XX H. pylori ORF hp3e11075orf3 protein.
DE
XX
XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
XX
XX Helicobacter pylori.
OS
XX

PH Key Location/Qualifiers
FT Misc-difference 261
FT /note= "the nucleotides encoding this amino acid
FT are not given in the specification"
XX
XX WO9737044-A1.
XX
XX 09-OCT-1997.
XX
XX 27-MAR-1997; 97WO-US05223.
XX
XX 06-DEC-1996; 96US-0761318.
PR 29-MAR-1996; 96US-0625811.
PR 02-APR-1996; 96US-0758731.
PR 25-OCT-1996; 96US-0736905.
PR 28-OCT-1996; 96US-0738859.
XX
XX (ASTR) ASTRA AB.
FA
XX Alm RA, Smith D;
PI WPI; 1997-503122/46.
DR N-PSDB; AAV24741.
XX
XX Helicobacter pylori nucleic acid sequences and encoded
XX polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX
XX Claim 14; Pages 558-559; 1145pp; English.
PS
XX This sequence is a H. pylori protein of unspecified function.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
XX
XX Sequence 261 AA;
SQ
Query Match 95.6%; Score 1199; DB 18; Length 261;
Best Local Similarity 94.8%; Pred. No. 5e-102;
Matches 236; Conservative 6; Mismatches 7; Indels 0; Gaps 0;
QY 1 MKGSLAIVLGSLLASGAFYTLADGMPAKOQHNTGESVELHFHYPIKQKQEPKNSHLV 60
DB 13 MKGSLAIVLGSLLASGTFYTLADGMPKQOHNNGESVELHFHYPIKQKQEPKNNHLV 72
QY 61 VLIEPKIEINKVIPESYQKEFEKSLFLQSLFLERKGYVSQFQKASEIPQDIKEKALLV 120
DB 73 VLIDPKIEANKVIPENYQKEFEKSLFLQSLNFLERKGYVSQFQKASEIPQDIKEKALLV 132
QY 121 LRMDGNVAILEDIVEESDALSEKVIDMSSGYLNLNFVEPKSEDIHHSFGIDVSKIKAVI 180
DB 133 LRMDGNVAILEDIVEESDALSEKVIDMSSGYLNLNFVEPKSEDIHHSFGIDVSKIKAVI 192
QY 181 ERVELARTNSGGFVPKTFVHRKETDHDQAIKIMNQAYHKVMVHITKELSKKHMEHYEK 240
DB 193 ERVELARTNSGGFVPKTFVHRKETDHDRAIKIMNQAYHKVMVHITKELSKKHMEHYEK 252

QY 241 VSSEMKKXK 249
 Db 253 VSSEMKKXK 261

RESULT 5
 ABUS1422
 ID ABUS1422 standard; Protein; 132 AA.
 XX AC
 XX ABUS1422;
 DT 07-MAY-2003 (first entry)
 DE Helicobacter pylori selected interacting domain (SID) protein #766.
 XX KW
 XX Protein-protein interaction; ulcer; selected interacting domain;
 XX KW SID.
 XX OS Helicobacter pylori.
 XX PN WO200266501-A2.
 XX PD 29-AUG-2002.
 XX PF 28-DEC-2001; 2001WO-EP15428.
 XX PR 02-JAN-2001; 2001US-259302P.
 XX PA (HYBR-) HYBRIGENICS.
 XX PA (INSP) INST PASTEUR.
 XX PI Legrain P, Rain J, Collard F, De Reuse H, Labigne A;
 XX WPI; 2002-674910/72.
 DR N-PSDB; ABX66167.

New complexes of protein-protein interactions in Helicobacter pylori,
 useful for identifying modulating compounds for treating or preventing
 ulcers in mammals -

Claim 6; Page 272; 642pp; English.

The invention describes a complex of protein-protein interactions in
 Helicobacter pylori selected from 421 complexes given in the
 specification. The complex of protein-protein interactions are useful
 for screening for agents which modulate the interaction of proteins.
 CC Modulating compounds which binds to a targeted bacterial protein may be
 CC used for treating or preventing ulcers in a human or animal. This
 CC is the amino acid sequence of a selected interacting domain (SID),
 CC identified via protein-protein interactions.
 CC Note: Where the patent number printed at the top of the pages in the
 CC specification has obscured areas of protein sequence, the indexer
 CC has replaced the residue with an X to represent an illegible residue.

QY Sequence 132 AA;
 Query Match 51.2%; Score 642; DB 23; Length 132;
 Best Local Similarity 100.0%; Pred. No. 3.4e-51;
 Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 SHLVLLIEPKIENKIPESYQKEFKSLFLQSSFLERKGYVSQFKDASEIPQDIK 116
 Db 1 SHLVLLIEPKIENKIPESYQKEFKSLFLQSSFLERKGYVSQFKDASEIPQDIK 60

QY 117 ALLVLRMDGNVAILEDIVEESDALSSEKVIDMSSGYLNLFNFBPKSEDIHFGDIVSKI 176
 Db 61 ALLVLRMDGNVAILEDIVEESDALSSEKVIDMSSGYLNLFNFBPKSEDIHFGDIVSKI 120

QY 177 KAVIERVELART 188
 Db 121 KAVIERVELART 132

RESULT 6
 AAY10951
 ID AAY10951 standard; Protein; 282 AA.
 XX AC
 XX AAY10951;
 DT 08-JUN-1999 (first entry)
 DE H. pylori ORF 04ee11108_3906963_fl_7 cell envelope protein.
 XX KW
 XX Vaccine; probe; diagnostic; ORF; cell envelope protein;
 KW secreted protein; cellular protein.
 XX OS Helicobacter pylori.
 XX PN WO9818323-A1.
 XX PD 07-MAY-1998.
 XX PF 28-OCT-1997; 97WO-US19575.
 XX PR 14-JUL-1997; 97US-0891928.
 XX PR 28-OCT-1996; 96US-0739150.
 XX PR 06-DEC-1996; 96US-0759739.
 XX PA (ASTR) ASTRA AB.
 XX PI Alm RA, Smith D;
 XX DR WPI; 1998-271811/24.
 XX DR N-PSDB; AAX30418.

Helicobacter pylori nucleic acids and proteins - used to develop
 products for the detection, prevention and treatment of H. pylori
 infections

Claims 27, 31; Page 151; 279pp; English.

Recombinant or substantially pure preparations of H. pylori polypeptides
 are disclosed, together with the nucleic acids encoding them. In all,
 73 ORFs are shown. The proteins are variously cell envelope proteins,
 secreted proteins or other cellular proteins. Vaccines containing the
 nucleic acids or proteins are claimed, as are probes containing at least
 8 nucleotides from the nucleic acid sequences. The vaccines are useful
 for treating or reducing the risk of H. pylori infections, and the
 probes can be used diagnostically for detecting the presence of
 Helicobacter in a sample. The products are also of use in screening
 for compounds having the ability to interfere with the H. pylori life
 cycle or to inhibit H. pylori infection.

QY Sequence 282 AA;
 Query Match 23.9%; Score 300; DB 19; Length 282;
 Best Local Similarity 29.7%; Pred. No. 3e-19;
 Matches 76; Conservative 57; Mismatches 107; Indels 16; Gaps 6;

QY 6 LAIVLGGILASGAPYATLADGMP--AKQO-----HNTGSEVSE---LHPHYPIKQKEP 54
 Db 18 VALGSSVLGICANPSEATYKKNDAKXQQPVQVQVQVQVQVQVQVQVQVQVQVQV 77

QY 55 KNSHLVLLIEPKIENKIPESYQKEFKSLFLQSSFLERKGYVSQFKDASEIPQDIK 114
 Db 78 QNHVWGILMPRIQVSDNL-KPYTDKFDALINQIQIPEKRGYQVLRFODEKALNVQDK 136

QY 115 EKALLVLRMDGNVAILEDI---VEESDALSSEKVIDMSSGYLNLFNFBPKSEDIHSGFI 171
 Db 137 KITSVLDLKGWGLDGLKKNLKNLKNLKNLKNLKNLKNLKNLKNLKNLKNLKNLKNL 196

QY 172 DVSKIKAVIERVELARTNSGGF-VPKTFVHRIKETDHDQAIRKINMQAYKVMVHITKEL 230
 Db 197 EVGTFQAITVYTTNNASGGFNSSKSVIHENLDKREDAIHKILNRYAVVMKAVTEL 256

QY 231 SKGMEHYEKVSSEMK 246

Db 257 TKENIAKYRDAIDRMK 272

RESULT 8
AAE15838
ID AAE15838 standard; Protein; 282 AA.
XX AC AAE15838;
XX DT 26-MAR-2002 (first entry)
XX DE Helicobacter pylori antigen Hpa44 #1.
XX KW Helicobacter pylori antigen; HpaA; Hpa44; pharmaceutical; vaccine;
XX KW infection; antibacterial.
XX OS Helicobacter pylori.
XX PN WO200183533-A1.
XX PD 08-NOV-2001.
XX PF 24-APR-2001; 2001WO-SE00872.
XX PR 29-APR-2000; 2000GB-0010371.
XX PA (ASTR) ASTRAZENECA AB.
XX PI Doig P, Pappo J;
XX DR WPI; 2002-082839/11.
XX DR N-PSDB; AAD25529.
XX PT Pharmaceutical product useful for treating and reducing the risk of
PT Helicobacter pylori infection in mammal, comprises HpaA and Hpa44
PT polypeptides or nucleotide sequences encoding the HpaA and Hpa44
PT polypeptides
XX Claim 7; Page 50-51; 61pp; English.
XX The invention relates to a pharmaceutical product, eg. a vaccine
CC composition or vaccine kit, which comprises Helicobacter pylori
CC antigens HpaA and Hpa44 polypeptides. The invention also concerns
CC a fusion polypeptide comprising both HpaA and Hpa44 amino acid
CC sequences and their corresponding nucleic acids. The fusion
CC polypeptides and their corresponding nucleic acids are useful in
CC the manufacture of vaccine, for administration to a mammal to
CC prevent and treat H.pylori infection. The present sequence is
CC Helicobacter pylori antigen Hpa44.
XX SQ Sequence 282 AA;
Query Match 23.9%; Score 300; DB 23; Length 282;
Best Local Similarity 29.7%; Pred. No. 3e-19;
Matches 76; Conservative 57; Mismatches 107; Indels 16; Gaps 6;
QY 6 LAIVGLSLLASGAFYALADGMP--AKQQ-----HNNTGESVE----LHFHYPIKQKQEP 54
DB 18 VALGLSVLLIGCAMNPSAETKKNDAKNQPVQVOTHERMTTSSHVTFDFNYPVHIVQAP 77
QY 55 KNSHLVVLLEPKIEINKVIPESYQKEPKSLFLQSLGFLERKGYSVSQFKDASEIPQDIK 114
DB 78 QNHVVGILMPRIQVSDNL-KPYIDKFDQALINQIOTIFEKRGYQVLRFODEKALNVQDK 136
QY 115 EKALLVLRMDGNVAILLEDI---VEESDALSEKVIDMSSGYLNLFNVEPKSEDIHSPGI 171
DB 137 KKFISVLDLKGWVGILEDLKNNLKDPSNPLDITLVQSSGVWVNFYEPESNRVVDFAV 196
QY 172 DVSKIKAVIERVELRTNSGGF-VPKTFVHRIKETHDQAIRKIMQOAHKVMVHITKEL 230
DB 197 EVGTFQAITTYTSTNNASGGFNSSKSVIHENLDKNREDIAHKILNRMIYAVVMKAVTEL 256
QY 231 SKKHEHYEKSSEMK 246

Db 257 TKENIAKYRDAIDRMK 272

RESULT 7
AAAY17214
ID AAAY17214 standard; Protein; 282 AA.
XX AC AAAY17214;
XX DT 03-AUG-1999 (first entry)
XX DE H. pylori outer membrane polypeptide.
XX KW Outer membrane polypeptide; OMP; vaccine; H. pylori infection; humoral;
XX KW cellular immune response.
XX OS Helicobacter pylori.
XX PN WO9921959-A2.
XX PD 06-MAY-1999.
XX PF 28-OCT-1998; 98WO-US22883.
XX PR 17-DEC-1997; 97US-0993001.
XX PR 28-OCT-1997; 97US-0959131.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX PI Alm RA, Ellis RW, Guild BC, Noonan BM, Smith D;
XX DR WPI; 1999-326698/27.
XX DR N-PSDB; AAX75833.
XX PT Cellular vaccine against Helicobacter pylori
XX Claim 7; Page 306-307; 352pp; English.
XX The invention relates to a vaccine for preventing or treating infections
CC by Helicobacter pylori. The vaccine contains at least one isolated
CC H. pylori polypeptide, or its fragments, in a carrier, where the
CC carrier is a Salmonella, Vibrio cholerae or Shigella vector containing a
CC nucleic acid encoding the H. pylori polypeptide. The vaccines induce
CC humoral and cellular immune responses. The vaccines are used to treat or
CC prevent infections by H. pylori. Sequences AAX75779 to AAX75837 represent
CC nucleic acid sequences encoding H. pylori outer membrane polypeptides
CC (OMPs) AAY17160 to AAY17218.
XX SQ Sequence 282 AA;
Query Match 23.9%; Score 300; DB 20; Length 282;
Best Local Similarity 29.7%; Pred. No. 3e-19;
Matches 76; Conservative 57; Mismatches 107; Indels 16; Gaps 6;
QY 6 LAIVGLSLLASGAFYALADGMP--AKQQ-----HNNTGESVE----LHFHYPIKQKQEP 54
DB 18 VALGLSVLLIGCAMNPSAETKKNDAKNQPVQVOTHERMTTSSHVTFDFNYPVHIVQAP 77
QY 55 KNSHLVVLLEPKIEINKVIPESYQKEPKSLFLQSLGFLERKGYSVSQFKDASEIPQDIK 114
DB 78 QNHVVGILMPRIQVSDNL-KPYIDKFDQALINQIOTIFEKRGYQVLRFODEKALNVQDK 136
QY 115 EKALLVLRMDGNVAILLEDI---VEESDALSEKVIDMSSGYLNLFNVEPKSEDIHSPGI 171
DB 137 KKFISVLDLKGWVGILEDLKNNLKDPSNPLDITLVQSSGVWVNFYEPESNRVVDFAV 196
QY 172 DVSKIKAVIERVELRTNSGGF-VPKTFVHRIKETHDQAIRKIMQOAHKVMVHITKEL 230
DB 197 EVGTFQAITTYTSTNNASGGFNSSKSVIHENLDKNREDIAHKILNRMIYAVVMKAVTEL 256
QY 231 SKKHEHYEKSSEMK 246

Db 257 TKENIAKYRDAIDRMK 272

RESULT 9

AAW20938

ID AAW20938 standard; protein; 285 AA.

AC AAW20938;

XX

XX

DT 21-JUL-1997 (first entry)

XX

XX

XX H. pylori secreted or periplasmic protein, 272el0351orf5.

XX

XX Cytoplasmic; vaccine; prevention; treatment; infection; identification;

KW binding compound; bacterium; life cycle; activator; inhibitor;

KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.

XX

OS Helicobacter pylori.

XX

XX WO9640893-A1.

XX

XX 19-DEC-1996.

XX

XX 06-JUN-1996; 96WO-US091122.

XX

XX 01-APR-1996; 96US-0630405.

PR 07-JUN-1995; 95US-0487032.

XX

XX (ASTR) ASTRA AB.

XX

XX

PI Berglindh OT, Smith D, Mellgaard BL;

XX

DR WPI; 1997-052306/05.

DR N-PSDB; AAT68191.

XX

XX Helicobacter pylori nucleic acid sequences and related

PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori

PT infection, and to detect Helicobacter

XX

PS Claim 72; Page 1333-1334; 1481pp; English.

XX

XX The present sequence is a H. pylori secreted or periplasmic protein.

CC The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds,

CC useful as potential H. pylori life cycle activators or inhibitors.

CC The genomic sequence of H. pylori (ATCC 55679) was determined from

CC overlapping contigs generated by mechanically shearing the bacterial

CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,

CC and the predicted coding regions defined by computer evaluation. To

CC identify likely H. pylori antigens for vaccine development, the amino

CC acid sequences predicted from various ORF were analysed for significant

CC homology to other known or exported membrane proteins. Having identified

CC and determined the sequences of interest, particular regions can be

CC isolated from H. pylori by PCR amplification for recombinant polypeptide

CC production, e.g. in E. coli hosts.

XX

SQ Sequence 285 AA;

Query Match 23.9%; Score 300; DB 18; Length 285;

Best Local Similarity 29.7%; Pred. No. 3e-19;

Matches 76; Conservative 57; Mismatches 107; Indels 16; Gaps 6;

QY 6 LAIVLGSLLASGAFYALADGMP--AKQO-----HNNTCESVE----LHFHYPIKQKEP 54

Db 21 VALGLSSVLIGCAMNPSAETKPDNDKQOPVQTHERTMTSSBHTPLDFNYPVHVQAP 80

QY 55 KNSHVVLLIEPKIEINKVPIESYQKEFKSLFLQLSSFLERKGYVSQPKDASEIPQDIK 114

Db 81 QNHVVGILMPRIQVSDNL-KPYIDFQDALINQIOTIFEKRGYQVLRFDQKALNVQDK 139

QY 115 EKALLVLRMDGNVAILEDI---VEESDALSSEKVIDMSSGYLNFVPEPKSEDIHSFGI 171

Db 140 KKIFSVLDLKGWVGILEDLKQMLKDPNSPNDLTLVDQSSGSWFWNFYFESNRVVDFAV 199

QY 172 DVSKIKAVIERVELRTNNGGF-VPKTFVHRIKETDHDQAIRKIMNQAYHKVWHITKEL 230

Db 200 EVGTQAITTYTSTNNASGGFNSKSVIHENLDKNREDAIHKILNRVAVVMKAVTEL 259

QY 231 SKHMEHYEKVSSENK 246

Db 260 TKENIAKYRDAIDRMK 275

RESULT 10

AAE15839

ID AAE15839 standard; Protein; 285 AA.

XX

AC AAE15839;

XX

DT 26-MAR-2002 (first entry)

XX

XX Helicobacter pylori antigen Hpa44 #2.

DE

XX Helicobacter pylori antigen; HpaA; Hpa44; pharmaceutical; vaccine;

KW infection; antibacterial.

XX

OS Helicobacter pylori.

XX

XX WO200183533-A1.

XX

XX 08-NOV-2001.

XX

XX 24-APR-2001; 2001WO-SE00872.

XX

XX 29-APR-2000; 2000GB-0010371.

XX

XX (ASTR) ASTRAZENECA AB.

PA

XX

PI Doig P, Pappo J;

XX

DR WPI; 2002-082839/11.

DR N-PSDB; AAD25530.

XX

XX Pharmaceutical product useful for treating and reducing the risk of

PT Helicobacter pylori infection in mammal, comprises HpaA and Hpa44

PT polypeptides or nucleotide sequences encoding the HpaA and Hpa44

XX polypeptides -

PS Claim 7; Page 51-52; 61pp; English.

XX

XX The invention relates to a pharmaceutical product, eg. a vaccine

CC composition or vaccine kit, which comprises Helicobacter pylori

CC antigens HpaA and Hpa44 polypeptides. The invention also concerns

CC a fusion polypeptide comprising both HpaA and Hpa44 amino acid

CC sequences and their corresponding nucleic acids. The fusion

CC polypeptides and their corresponding nucleic acids are useful in

CC the manufacture of vaccine, for administration to a mammal to

CC prevent and treat H. pylori infection. The present sequence is

CC Helicobacter pylori antigen Hpa44.

XX

SQ Sequence 285 AA;

Query Match 23.9%; Score 300; DB 23; Length 285;

Best Local Similarity 29.7%; Pred. No. 3e-19;

Matches 76; Conservative 57; Mismatches 107; Indels 16; Gaps 6;

QY 6 LAIVLGSLLASGAFYALADGMP--AKQO-----HNNTCESVE----LHFHYPIKQKEP 54

Db 21 VALGLSSVLIGCAMNPSAETKPDNDKQOPVQTHERTMTSSBHTPLDFNYPVHVQAP 80

QY 55 KNSHVVLLIEPKIEINKVPIESYQKEFKSLFLQLSSFLERKGYVSQPKDASEIPQDIK 114

Db 81 QNHVVGILMPRIQVSDNL-KPYIDFQDALINQIOTIFEKRGYQVLRFDQKALNVQDK 139

QY 115 EKALLVLRMDGNVAILEDI---VEESDALSSEKVIDMSSGYLNFVPEPKSEDIHSFGI 171

140 KVFVLDLKGWVGILEDLKMLKDPNSFNLDTLVDQSSGVSWMFVPEPESNRVVDFAV 199
 172 DYSKIKAVIERVELRTNSGGF-VPKTFVRIKETHDQAIKRNQAYHKVMVHITKEL 230
 200 EVGTQAITTYTSTNNASGGFNSKSVIHENLDKREDALHKILNRNAYVVMKAVTEL 259
 231 SKKHMEHYEKVSSEMK 246
 260 TRENIAKYRDAIDRMK 275

RESULT 11
 AAB21214
 ID AAB21214 standard; Protein; 254 AA.
 XX
 AC AAB21214;
 XX
 DT 18-DEC-2000 (first entry)
 XX
 DE Helicobacter pylori partial Hpa44.
 XX
 KW Helicobacter pylori; Hpa44; Pichia pastoris; infection; vaccine;
 XX protein expression.
 KW
 OS Helicobacter pylori.
 XX
 PN WO200049044-A1.
 XX
 PD 24-AUG-2000.
 XX
 PF 18-FEB-2000; 2000WO-SE00340.
 XX
 PR 19-FEB-1999; 99US-0253290.
 XX
 PA (ASTR) ASTRAZENECA AB.
 XX
 PI Alm R, Wengender P, Zen K;
 XX
 DR WPI; 2000-515101/46.
 XX
 DR N-PSDB; AAA/4760.
 XX
 PT Producing Helicobacter polypeptides useful for vaccines for treatment
 of infection, comprises transformation of Helicobacter nucleotide
 sequence into Pichia yeast and culturing -
 PT
 PS Disclosure; Page 28-29; 33pp; English.
 XX
 CC The present sequence is the Hpa44 polypeptide from Helicobacter pylori.
 CC The nucleotide sequence encoding this polypeptide may be cloned into a
 CC yeast expression vector and used to transform Pichia pastoris. The
 CC transformed yeast is then cultured and Hpa44 polypeptide is expressed.
 CC The purified polypeptide may be used in a vaccine for the treatment
 CC and/or prophylaxis of a pre-existing Helicobacter infection. It may also
 CC be used as a target in drug discovery, for example in an assay to
 CC identify an agonist or antagonist of the polypeptide. Significant amounts
 CC of the Helicobacter polypeptide can be obtained by this method. The
 CC polypeptide is not post-translationally modified by the Pichia yeast,
 CC which is useful when the polypeptide is to be used in a vaccine.
 CC
 SQ Sequence 254 AA;
 Query Match 23.7%; Score 297.5; DB 21; Length 254;
 Best Local Similarity 30.4%; Pred. No. 4.3e-19;
 Matches 68; Conservative 53; Mismatches 94; Indels 9; Gaps 4;

31 QCHNNTGESVE-----LHFHYPIKGQEPKNSHLVLEPKIEINKVIPESYQKEPKSLF 86
 22 QTHRMVTSSEHVTPLDFNYPVHVQAPQNHVVGILMPRIQVSDNL-KPYIDKFODALI 80
 87 LQLSSFLERKGVSYQFKDASEIPQDIKEKALLVLRMDGNVAILEDI---VEESDALSEE 143
 81 NQIQTFEKGQVLRVQDEKALNVQDKKIFSVLDLKGWVGILEDLKYNLKDPSNPDL 140

144 KVIDMSSGYNLNFPVPEKSEDIHSFGIDVSKIKAVIERVELRTNSGGF-VPKTFVHRI 202
 141 TLVDQSSGVSWMFVPEPESNRVVDFAVEVGTQAITTYTSTNNASGGFNSKSVIHEN 200
 203 KETHDQAIKRNQAYHKVMVHITKELSKKHMEHYEKVSSEMK 246
 201 LDKNREDALHKILNRNAYVVMKAVTELTKENAKYRDAIDRMK 244

RESULT 12
 AAE15841
 ID AAE15841 standard; Protein; 254 AA.
 XX
 AC AAE15841;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Helicobacter pylori antigen Hpa44 #4.
 XX
 KW Helicobacter pylori antigen; Hpa44; Hpa44; pharmaceutical; vaccine;
 XX infection; antibacterial.
 KW
 OS Helicobacter pylori.
 XX
 PN WO200183533-A1.
 XX
 PD 08-NOV-2001.
 XX
 PF 24-APR-2001; 2001WO-SE00872.
 XX
 PR 29-APR-2000; 2000GB-0010371.
 XX
 PA (ASTR) ASTRAZENECA AB.
 XX
 PI Doig P, Pappo J;
 XX
 DR WPI; 2002-082839/11.
 XX
 DR N-PSDB; AAD25532.
 XX
 PT Pharmaceutical product useful for treating and reducing the risk of
 Helicobacter pylori infection in mammal, comprises Hpa44 and Hpa44
 polypeptides or nucleotide sequences encoding the Hpa44 and Hpa44
 polypeptides -
 PT
 PS Claim 7; Page 56; 61pp; English.
 XX
 CC The invention relates to a pharmaceutical product, eg, a vaccine
 CC composition or vaccine kit, which comprises Helicobacter pylori
 CC antigens Hpa44 and Hpa44 polypeptides. The invention also concerns
 CC a fusion polypeptide comprising both Hpa44 and Hpa44 amino acid
 CC sequences and their corresponding nucleic acids. The fusion
 CC polypeptides and their corresponding nucleic acids are useful in
 CC the manufacture of vaccine, for administration to a mammal to
 CC prevent and treat H.pylori infection. The present sequence is
 CC Helicobacter pylori antigen Hpa44.
 CC
 SQ Sequence 254 AA;
 Query Match 23.7%; Score 297.5; DB 23; Length 254;
 Best Local Similarity 30.4%; Pred. No. 4.3e-19;
 Matches 68; Conservative 53; Mismatches 94; Indels 9; Gaps 4;

31 QCHNNTGESVE-----LHFHYPIKGQEPKNSHLVLEPKIEINKVIPESYQKEPKSLF 86
 22 QTHRMVTSSEHVTPLDFNYPVHVQAPQNHVVGILMPRIQVSDNL-KPYIDKFODALI 80
 87 LQLSSFLERKGVSYQFKDASEIPQDIKEKALLVLRMDGNVAILEDI---VEESDALSEE 143
 81 NQIQTFEKGQVLRVQDEKALNVQDKKIFSVLDLKGWVGILEDLKYNLKDPSNPDL 140

Db 141 TLVDQSSGVWVNFYEPESNRVVDHFAVEVGTFOAITTYTSTNNASGGFNSSKSVIHEN 200
 QY 203 KETDHDQAIRKIMNOAYHKVWVHITKELSKGHMEHYKVSSEMK 246
 Db 201 LDKNREDAIHKILNRMYAVVMKKAVTELTKENIAYRDAIDRMK 244

RESULT 13
 AAB21215
 ID AAB21215 standard; Protein; 396 AA.
 AC AAB21215;
 XX
 DT 18-DEC-2000 (first entry)
 XX
 DE cyt/Hpa44/HF fusion protein.
 XX
 KW cyt/Hpa44/HF; fusion protein; Helicobacter pylori; Hpa44;
 KW cytochrome b562; flag epitope tag; Pichia pastoris; infection; vaccine;
 KW protein expression.
 XX
 OS Escherichia coli.
 OS Helicobacter pylori.
 OS Synthetic.
 XX
 PN WO200049044-A1.
 XX
 PD 24-AUG-2000.
 XX
 PF 18-FEB-2000; 2000WO-SE00340.
 XX
 PR 19-FEB-1999; 99US-0253290.
 XX
 PA (ASTR) ASTRAZENECA AB.
 XX
 PI Alm R, Wengender P, Zen K;
 XX
 DR WPI; 2000-515101/46.
 XX
 CC Producing Helicobacter polypeptides useful for vaccines for treatment
 FT of infection, comprises transformation of Helicobacter nucleotide
 FT sequence into Pichia yeast and culturing -
 XX
 PS Disclosure; Page 29-30; 33pp; English.
 XX

The present sequence is a fusion protein comprising Escherichia coli
 CC cytochrome b562, the Hpa44 polypeptide from Helicobacter pylori,
 CC six tandem histidine residues and the flag epitope tag. The nucleotide
 CC sequence encoding this protein was cloned into a yeast expression vector
 CC and used to transform Pichia pastoris. The transformed yeast was then
 CC cultured and the fusion protein was expressed. The FLAG tag is used to
 CC purify the protein and cytochrome b562 acts as an expression reporter.
 CC The Hpa44 polypeptide may be used in a vaccine for the treatment
 CC and/or prophylaxis of a pre-existing Helicobacter infection. It may also
 CC be used as a target in drug discovery, for example in an assay to
 CC identify an agonist or antagonist of the polypeptide. Significant amounts
 CC of the Helicobacter polypeptide can be obtained by this method. The
 CC polypeptide is not post-translationally modified by the Pichia yeast,
 CC which is useful when the polypeptide is to be used in a vaccine.
 XX

SQ Sequence 396 AA;
 Query Match 23.7%; Score 297.5; DB 21; Length 396;
 Best Local Similarity 30.4%; Pred. No. 8.2e-19;
 Matches 68; Conservative 53; Mismatches 94; Indels 9; Gaps 4;

QY 31 QOHNNTGESVE---LHFHYPIKQEPKNSHLVLIPIEKIENKVPESYQKEFEKSLF 86
 Db 139 QTHERTMTTSSEHTPLDYNFVHIVQAPQNHVVGILMPRIQVSDNL-RPYIDKFDALI 197
 QY 87 LQLSSFLERKGVSVQFKDASEIPQDIKEKALLVLRMDGNVAILEDI---VSESDALSEE 143
 Db 198 NOQTIFERKGVQLRFQDEKALNVQDKKIFSVLDLKGWVGLLEDLQNLKNDPNSPLD 257

QY 144 KVDMSSGYLNNFVEPKSEDIHSGFDIVSKIKAVIERVELRRTNSSGF-VPKTLFVHRI 202
 Db 258 TLVDQSSGVWVNFYEPESNRVVDHFAVEVGTFOAITTYTSTNNASGGFNSSKSVIHEN 317
 QY 203 KETDHDQAIRKIMNOAYHKVWVHITKELSKGHMEHYKVSSEMK 246
 Db 318 LDKNREDAIHKILNRMYAVVMKKAVTELTKENIAYRDAIDRMK 361

RESULT 14
 AAO20108
 ID AAO20108 standard; Protein; 337 AA.
 AC AAO20108;
 XX
 DT 06-AUG-2002 (first entry)
 XX
 DE Protein #1 encoded by 1516nt DNA related to Helicobacter pylori vaccine.
 XX
 KW Vaccine; Helicobacter pylori.
 XX
 OS Unidentified.
 XX
 PN KR99076437-A.
 XX
 PD 15-OCT-1999.
 XX
 PF 31-MAR-1998; 98KR-0011379.
 XX
 PR 31-MAR-1998; 98KR-0011379.
 XX
 PA (DAEW-) DAEWOONG PHARM CO LTD.
 XX
 PI Kim BO, Lee BG, Yoon SW, Park SG, Yoo YH, Pyo SN, Choi DJ;
 PI Chung ;
 XX
 DR WPI; 2000-585436/55.
 DR N-PSDB; AAK99547.
 XX
 PT Vaccine for prevention and remedy of diseases concerning Helicobacter
 PT pylori -
 XX
 PS Disclosure; Fig 1; 12pp; Korean.
 XX

The invention relates to a vaccine for the prevention and remedy of
 CC diseases concerning Helicobacter pylori. This sequence represents a
 CC protein region encoded by a 1516 nucleotide DNA sequence relating to a
 CC Helicobacter pylori vaccine of the invention.
 XX

SQ Sequence 337 AA;
 Query Match 16.3%; Score 204; DB 21; Length 337;
 Best Local Similarity 24.5%; Pred. No. 2.6e-10;
 Matches 69; Conservative 64; Mismatches 103; Indels 46; Gaps 11;

QY 1 MKKGSLLAIVIGSLIAS-----GAFVTLADGMPAKQOHNNTGESVELHFHY 46
 Db 1 MKKTAIAIATAVA--LAGFASVAQAAMKCLLGASVALLVGC---SPHIIETNEVALKLY 55
 QY 47 PIKGOEPKNSHLVLIPIEKIENKVPESYQKEFEKSLFOLSSFLERKGVSV---SQ 102
 Db 56 HPASEKQVALDEKILLRPAFOYSDNIAKEYENKFNQTKALKVEQILONGYKIVSDSS 115
 QY 103 FKDASEIPQDIKEKALLVLRMDGNVAILEDI---IVEESD-----ALSEEKVIDMSG 151
 Db 116 DKDDFSFAQ--KKEGYLAVAMNGEIVLRPDPKRTIQKSEFGLLFSTGLDKMEGVLIAP 173
 QY 152 YLNNFVEPKSEDIHSGFDIVSKIKAVIERVELRRT---NSGGFVPKTFVHRIKTDH- 207
 Db 174 FIKTILPMSGESLDSFTWDLSELD--IOEKFLKTHSSHSGGLVST-----WVKGTDS 227
 QY 208 DQAIRKIMNOAYHKVWVHITKELSKGHMEHYKVSSEMKKKR 249

228 NDAIKSALNKFANIMQEIDKKLTQKNLESYQDKAKELGKR 269

RESULT 15

AAW74466
ID AAW74466 standard; Protein; 461 AA.

XX
AC

XX
DT 25-MAR-2003 (updated)

DT 18-MAY-1999 (first entry)

XX
DE Adhesin/V.cholerae toxin A2 and B subunit fusion protein.

XX Fusion gene; H. pylori infection; adhesin gene; V. cholerae toxin;
KW A2 subunit; B subunit; diagnosis; vaccine; gastritis; gastric ulcer;
KW duodenal ulcer.

Chimeric - Helicobacter pylori.

OS Chimeric - *Vibrio cholerae*.

XX
PN W09853082-A1.

XX
PD
26-NOV-1998.

21-MAY-1997: 97WQ-KR000091

21-MAY-1997: 97WQ-KR000091

(DAEW-) DAEWONG PHARM CO LTD.

[illegible]

XX
XX
XX

DR N-PSDB; AAX21955.

Fusion gene consisting of the Helicobacter pylori adhesin gene ligated to the A2 and B subunit gene of Vibrio cholerae toxin - useful in the diagnosis and treatment of H. pylori

PS
XX
Claim 2; Page 20-21; 32pp; English.

This sequence represents the fusion protein of the invention, between the H. pylori adhesin gene, and the V. cholerae toxin A2 and B subunit genes. This protein is used in the diagnosis of and as a vaccine against H. pylori, which is the main cause of gastritis and gastric/duodenal ulcers. It may also be used in the production of an anti-H. pylori antibody. Vectors containing the DNA sequence of an anti-H. pylori containing it form the basis of a process to produce the protein, which consists of culturing the microorganism and obtaining the protein. The protein is an effective vaccine against H. pylori due to its excellent immunogenicity for H. pylori, stability within the stomach environment, and its ability to penetrate through the stomach mucous membrane to stimulate siqa production.

CC Note: This sequence was indexed from WO9853082, which is the first
CC major country equivalent to KR97059278.

CC (Updated on 25-MAR-2003 to correct PF field.)
CC (Updated on 25-MAR-2003 to correct DB field.)
CC (Updated on 25-MAR-2003 to correct DB field.)

XX	
SC	Sequence 461 AA:

Query Match

Exact Match 10.5%; Score 204; Length 401;
Best Local Similarity 24.5%; Pred. No. 4.1e-10;
Matches 59; Conservative 54; Mismatches 103; Indels 4

007, conservative	04, mismatches	103
Matches		
007		
1 MRRGSIATVTCSTLAS		CAEVTAIAT

[illegible]

1 MKKXIMIAVA--TAFGASVQAAWAKKCLLEGGASVVALLVGC---SPRIIEINEVAUKNYI 53

4/ PINGQEFANSHLVLLIEFKLEINNVIPESIQEFERSDFQSSFLKRGYSV-----SQ

DD 30 HFASENVQALDEKILDEUKRPAQISDNIAHE:FNKFNQIADURVEQILQNGKIRVISVDSS 111

QY 99 SV-----SOFKDASEIPQDIKEKALLVLRMDGNVAILED-----IYVESD-----ALSEE 143
DB 97 KVISVDSDDKDFSAQ--KRGYLVAMNGEIVLRPDKRTQKSEPGLLFSTGLDKM 154
QY 144 KVIDMSSGYLNLNVEPKSEDIHSGFIDVSKIKAVIERVELRRT---NSGGFVPKTFVH 200
DB 155 EGVLIIPAGFKVITLPEMSEGLDSFTMDLSELD--IQEKFLKTHSHSGGLVST---- 208
QY 201 RIKETDH--DOAIRKIMNQAYHKVMVHTKELSKKMEHYKVSSEMKKKR 249
DB 209 MVKGTDSNDAIKRALNKIFANIMOEDKLTQKNLESYQKDAKELKGR 258
RESULT 2
US-08-669-560-2
; Sequence 2, Application US/08669560
; Patent No. 6025164
; GENERAL INFORMATION:
; APPLICANT: Bolin, Ingrid
; APPLICANT: Svennerholm, Ann-Mari
; TITLE OF INVENTION: Bacterial Antigens and Vaccine
; TITLE OF INVENTION: Compositions
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version
; SOFTWARE: #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,560
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Sturner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-0220
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8200
; TELEFAX: (212)354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 260 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-669-560-2
Query Match 15.9%; Score 199.5; DB 3; Length 260;
Best Local Similarity 24.8%; Pred.No.6.1e-12;
Matches 57; Conservative 65; Mismatches 79; Indels 29; Gaps 10;
QY 39 SVLFHYPYIKGOEPKNSHLVLIPEKIEINKVIPESYQKEFKSLFLQLSFLERKGY 98
DB 39 ALKLNHY-PASEKQVALDER-IULLRPAQFSDNIAYENKPKNQTKALKVEQILNQY 96
QY 99 SV-----SOFKDASEIPQDIKEKALLVLRMDGNVAILED-----IYVESD-----ALSEE 143
DB 97 KVISVDSDDKDFSAQ--KRGYLVAMNGEIVLRPDKRTQKSEPGLLFSTGLDKM 154
QY 144 KVIDMSSGYLNLNVEPKSEDIHSGFIDVSKIKAVIERVELRRT---NSGGFVPKTFVH 200
DB 155 EGVLIIPAGFKVITLPEMSEGLDSFTMDLSELD--IQEKFLKTHSHSGGLVST---- 208
QY 201 RIKETDH--DOAIRKIMNQAYHKVMVHTKELSKKMEHYKVSSEMKKKR 249

DB 209 MVKGTDSNDAIKRALNKIFANIMOEDKLTQKNLESYQKDAKELKGR 258
RESULT 3
US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Polypeptide
US-08-973-462-8
Query Match 8.7%; Score 108.5; DB 3; Length 1786;
Best Local Similarity 26.0%; Pred.No.0.12;
Matches 50; Conservative 35; Mismatches 90; Indels 17; Gaps 7;
QY 62 LIEPKIEINKVIPESYQKEFKSLFLQLSFLERKGYSVSQFDASIPQDIKEKALLVL 121
DB 1355 ILEKKEIEKDHFEKFEFEAEIEIKDLADILKEVSSLEVEEEKLEEV-HELKEVEHII 1413
QY 122 RMDGNV-ALLEDIVESDALSSEKVIDMSSGYLNLNVEPKS-EDIHSGFIDVSKIKAV 179
DB 1414 SGDAHIKGLEEDDLEEVDDL-KGSILDLKGMELGMDKESLEDDVTKLGERVESLKDV 1472
QY 180 IERV-----ELURRTNSGGFVPKTFVHRIKETDHDQAIRKIMNQAYHKVMVHTKELSKKH 234
DB 1473 LSSALGMDEEQMTRKVAQRPKLEEVLLKEEVKEPKKITKK---KVRFDIKDKPEKD- 1528
QY 235 MEHYEKVSSEMK 246
DB 1529 ----EIVEVEMK 1536
RESULT 4
US-09-914-259-11
; Sequence 11, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-914-259-11
Query Match 8.6%; Score 107.5; DB 4; Length 3878;
Best Local Similarity 20.7%; Pred.No.0.46;
Matches 52; Conservative 47; Mismatches 101; Indels 51; Gaps 10;

Db 595 EQGVKAREMANEFKSIQEDLVYERN-RVLEIDDAEN--RDFKALAKDVEM----- 646
Qy 129 ILEDIVESDALSBEKVIDSSGYLNINFPVKSEDIHSGFDVSKI-----KAVI--- 180
Db 647 ----FVNEEKVLTKSRVVEVI--YONLSF-----QFNKDVACVNFKDQAVVTEL 690
Qy 181 -----ERVELRRTSGG-----FVPKTFVHRI-----KETDHOAIRKIMNQ----- 217
Db 691 LEQEKQLALNRKMQSAYYINIFVQVFLKADSCWLEQVDYLOQLKASVNRQGRQNRN 750
Qy 218 ---AYHKVMVHITKELSKKMEHYEKSSEMKKR 248
Db 751 AIPFHRVA-----LDSFENVTRNIKKR 773

RESULT 7
US-09-134-001C-4284
; Sequence 4284, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: LYNN DOUCETTE-STAMM ET AL
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1997-11-08 964
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08 779
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4284
; LENGTH: 457
; TYPE: PR1
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4284

Query Match 7.7%; Score 96; DB 4; Length 457;
Best Local Similarity 24.5%; Pred. No. 0.28;
Matches 47; Conservative 41; Mismatches 64; Indels 40; Gaps 12;

Qy 43 HEHYPIKQEPKPNKSHLVLPKINVKIPESYQKEFKSLFLQSLF-LER--KGYS 99
Db 280 HYHPIGLTNNMNAIAIAIGHGLNETI---IQNNIHN---VOLTAMRHERHSSNN 333
Qy 100 VSOFKDA-SBIPQDIKEKALLVLRMDG-NVAILEDIVE--ESDALSEEKVIDMSSGYLNL 155
Db 334 VTINDAYNASPTSMKAAIDTLGSMKGRKILILADVLGLPNSQLMHKQVGE----- 385
Qy 156 NFVEPKSEDIHSGFDVSKIKAVIERVELRRTNSGGFVPKTFVHRIKETDH-DQAIRKI 214
Db 386 -YLDKKNIDVLYTFGEASYIH-----DSG-----KIFVKEAKYFDNKEQLIOTL 429
Qy 215 MNQA--YHKVMV 224
Db 430 ISQVKPEDKVLV 441

RESULT 8
US-08-021-601-2
; Sequence 2, Application US/08021601
; Patent No. 5591631
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Nichols, Peter J.
; APPLICANT: Arora, Naveen
; APPLICANT: Singh, Yogendra
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,601
; FILING DATE: 19930212
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 776 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-021-601-2

Query Match 7.5%; Score 94.5; DB 1; Length 776;
Best Local Similarity 20.9%; Pred. No. 0.86;
Matches 46; Conservative 49; Mismatches 80; Indels 45; Gaps 10;

Qy 30 KOCHNTGESVELHFPYIKQBPKNHVLVLPKIEINKVIPESYQKEF-EKSLFLQ 88
Db 207 EONSNEVOEVAKAFAYI---EPQHRDLQLYAPE-----AFNYMDKFNEQENLS 255
Qy 89 LSSFLERKGYVSQFKDAEIPQDIKEKALLVLRMDGNVAILEDIVEESDALSEKVIDM 148
Db 256 LEELKQOR--MLSYKXWKIKQHYQ-----HMSDSLSEE----- 288
Qy 149 SSGYL-NLNP-VEPKSEDIHSGFDVSKIKAVIERVELRRTNSGGFVPKTFVHRIKETD 206
Db 289 GRGLKXKLQPIEPKDDIHSLS---QEEKELKRIQDSDFLSTEEKEFLKKL-QID 344
Qy 207 HDQAIRKIMQAVHKVMVHITKELSKKMEHYEKSSEM 246
Db 345 IRDSLSEEEKELNRIQVDSNPLSEKEKEFLKKLIDIQ 384

RESULT 9
US-08-082-949B-2
; Sequence 2, Application US/08082849B
; Patent No. 5677274
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Arora, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
; TITLE OF INVENTION: Related Methods
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 776 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-2
US-08-082-849B-2

Query Match 7.5%; Score 94.5; DB 1; Length 776;
Best Local Similarity 20.9%; Pred. No. 0.86;
Matches 46; Conservative 49; Mismatches 80; Indels 45; Gaps 10;

QY 30 KQHNNTGSEVLFHYPYIKGQEPKNSHLVLIIEPKIEINKVIPESYQKEF-EKSLFLQ 88
Db 207 EQNSNEVQEVFAKAFAYI-----EPQHRDVLQLYAPE-----AFNYMKENEQEIINLS 255
QY 89 LSSFLERKGYVSQFQDASEIPQDIKEKALLVLRMDGNVAILEDIVESDALSSEKVIDM 148
Db 256 LEELKQQR--MLSRYEKWEKIKQHYQ-----HWSDSLSEE----- 288
QY 149 SSGYL-NLNF-VEPKSEDIHSFGIDVSKIKAVIERVELRRNNSGGFVPKTFVHRIKETD 206
Db 289 GRGLLKLQIPIEPKDDIHSLS---QEEKELKRIQIDSSDFLSTERKEFLKKL-QID 344
QY 207 HDQAIRKINQAYHKVMVHITKELSKHMEHYEKSSEM 246
Db 345 IRDSLSEEEKELNRIQVDSNPLSEKEKEFLKKLQID 384

RESULT 10
PCT-US94-01624-2
; Sequence 2, Application PC/TUS9401624
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Arora, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESS: TOWNSEND and TOWNSEND KHOURIE and CREW
; STREET: Stewart Street Tower, 20th Floor, One Market
; STREET: Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01624
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 776 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-2
US-08-082-849B-2

Query Match 7.5%; Score 94.5; DB 5; Length 776;
Best Local Similarity 20.9%; Pred. No. 0.86;
Matches 46; Conservative 49; Mismatches 80; Indels 45; Gaps 10;

QY 30 KQHNNTGSEVLFHYPYIKGQEPKNSHLVLIIEPKIEINKVIPESYQKEF-EKSLFLQ 88
Db 207 EQNSNEVQEVFAKAFAYI-----EPQHRDVLQLYAPE-----AFNYMKENEQEIINLS 255
QY 89 LSSFLERKGYVSQFQDASEIPQDIKEKALLVLRMDGNVAILEDIVESDALSSEKVIDM 148
Db 256 LEELKQQR--MLSRYEKWEKIKQHYQ-----HWSDSLSEE----- 288
QY 149 SSGYL-NLNF-VEPKSEDIHSFGIDVSKIKAVIERVELRRNNSGGFVPKTFVHRIKETD 206
Db 289 GRGLLKLQIPIEPKDDIHSLS---QEEKELKRIQIDSSDFLSTERKEFLKKL-QID 344
QY 207 HDQAIRKINQAYHKVMVHITKELSKHMEHYEKSSEM 246
Db 345 IRDSLSEEEKELNRIQVDSNPLSEKEKEFLKKLQID 384

RESULT 11
US-09-107-532A-6725
; Sequence 6725, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FACIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Denéke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
```

```

; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6725:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 413 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...413
; SEQUENCE DESCRIPTION: SEQ ID NO: 6725:
US-09-107-532A-6725

```

```

Query Match      7.5%; Score 94; DB 4; Length 413;
Best Local Similarity 24.8%; Pred. No. 0.38;
Matches 65; Conservative 39; Mismatches 76; Indels 82; Gaps 17;

QY 2 KKG-SLAIVLGSLLASGAFYALADGMPAKQOHNNTGESVELHFHYPIKQ-----51
Db 37 EKG-SKEIVDIALSEA-----AL--VSAYKRHYGQAQNVVEFDQK-KGKIHYAVKEV 86
QY 52 -QEPKNSHLVV-----LIEPKIEIN-----KVIPESYQK---EFKSLFLQLSSFLERK 96
Db 87 TEEVMDQLEVSLKDALLINPAVEIGDTIRFEVTPKDFGRIAACTAKQVILQVRRAER- 145
QY 97 GYSVQPKDASEIPQDIKEKALLVLRMDGNVAILEDIVESDALSEBEKVIDMSSGYLNLN 156
Db 146 ---TIIYNESAYEKD-----IMQGIVERQD---KRYI-----YVNLG 177
QY 157 FVE---PKSEDIHSGIDVSKIKAVIERVELARTNSGGFVPTFVHRIKETDHDQAIK 213
Db 178 KIEAVLSKQOMPNEFFQPHDRIKVYSRVE---NTSKG---PQVFVSR-----SHPDLRR 228
QY 214 IMNQAYHKV-----MVHITKE 229
Db 229 LFEQEVPEVYDGIIVEIVSIARE 250

```

```

RESULT 12
US-09-564-805-234
; Sequence 234, Application US/09564805
; Patent No. 633403
; GENERAL INFORMATION:
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng David H.F.
; APPLICANT: Simard, Jacques
; APPLICANT: Rommens, Johanna M.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: Chromosome 17p-Linked Prostate Cancer Susceptibility
; FILE REFERENCE: 2318-258
; CURRENT APPLICATION NUMBER: US/09/564,805
; CURRENT FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/107,468
; PRIOR FILING DATE: 1998-11-06
; PRIOR APPLICATION NUMBER: 09/434,382
; PRIOR FILING DATE: 1999-11-05
; NUMBER OF SEQ ID NOS: 240
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 234
; LENGTH: 693
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-564-805-234

```

```

Query Match      7.5%; Score 94; DB 4; Length 693;
Best Local Similarity 23.1%; Pred. No. 0.82;
Matches 37; Conservative 26; Mismatches 61; Indels 36; Gaps 6;

QY 4 GSLAIV---LGSLLASGAFYALADGMPAKQOHNNTGESVELHFHYPIKQEP-----K 55

```

```

Db 528 GAGGVIKHLEKIFESVEFSTDEESGLPALKVHERVTVKQSEKHISLOWSSDPISDMVS 587
QY 56 NSHLVVLPKPKIEINKVIPESYQ-----KEFEKSLFLQLSSFLERKGYSGVQPKDA 106
Db 588 DSIVALILNISREVPKIVMEBEDAVKSPENGKKVKVIYALL-----VSLFGDV 637
QY 107 SEIPQDIKEKALLVLRMDGNVAILE---DIVEESDALSE 142
Db 638 K-----LGENGKLVIRVDGNVAQLDKESGEVESEHSGLKE 672

```

```

RESULT 13
US-09-198-452A-405
; Sequence 405, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, preve
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 405
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; NAME/KEY: SITE
; LOCATION: 1...325
; OTHER INFORMATION: Xaa=unknown or other
US-09-198-452A-405

```

```

Query Match      7.5%; Score 93.5; DB 4; Length 325;
Best Local Similarity 20.8%; Pred. No. 0.3;
Matches 40; Conservative 33; Mismatches 54; Indels 65; Gaps 6;

QY 76 SYQKEFEKSLFLQLSSFLERK-----YSVQPKDASEIPQDIKEKALLVLRMDGNVAIL 130
Db 34 SFEAEFPSPYLPALENFQGTGEIELLYSSPK-----AKEKRVLLGLCKNEELT 83
QY 131 BDIVEESDAL-----SEEKVIDMSSGYLNLFVEPKSD 164
Db 84 SDVVFQYVILTRVLRKAKSTVNIILPTISEURLSAEEFLVGLSSGILSLNDYPRNK 143
QY 165 IHSFGIDVSKIKAVIERVELRRTNSGGFVPTFVHRIKETDHDQAIKIMNQAYHKVMV 224
Db 144 VDRNLETPLSKVTVI-----GIVPKM-----ADAIFRK---EAAIFEGV 179
QY 225 HITKELSKHME 236
Db 180 YLRDLVNRNAD 191

```

```

RESULT 14
US-09-702-705-1816
; Sequence 1816, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C14

```

57 DDIELRGAAGAAAAAPPPPIEECEPEDLPEK-----FDGNPDMLAPFMAQCOIFMEKSTRD 110
148 MSSGYLNLNFE-----PKSEDIHSGIDVSKIKAVIERVELRRTNSGG 192
111 FSVDRVRVCFTSMWTGGAARWASAKLERSHYLMHNYPAFMMEMKHVFPEDPORRE-----165
193 FVPKTFVHRIKE-----TDHQAIRKI-----MNOAYHKVMVHTTKELSKKXME 236
166 -VAKRKIRLRQGMGVIDYSNAFQMIADLDWNEPALIDQYHEGLSDHIQEELS--HLE 222
237 HYEKVSS 243
223 VAKLSLA 229

Search completed: December 10, 2003, 18:38:36
Job time : 17.644 secs

Db Qy
Qy Db
Qy Db
Qy Db
Qy Db

Query Match
Best Local Similarity 19.8%; Pred.No. 0.34; 91; Indels 62; Gaps 9;
Matches 49; Conservative 45; Mismatches 91; Indels 62; Gaps 9;

Qy 31 QOHNNTGESVLFHYPYKQPKNSHLVLIIEPKIEINKVIPESYQKEFEKSLFLQLS 90
Db 11 ESINNLRKYM-----KQSENNNLQSQVKLTENTTLREQVEPTPEDE-----D 56
Qy 91 SFLEKRGYSVSQFKD---ASEIPQDIKEKALLVLRMDGNVAILEDIVESDALSEKVID 147
Db 57 DDIELRGAAGAAAAAPPPPIEECEPEDLPEK-----FDGNPDMLAPFMAQCOIFMEKSTRD 110
Qy 148 MSSGYLNLNFE-----PKSEDIHSGIDVSKIKAVIERVELRRTNSGG 192
Db 111 FSVDRVRVCFTSMWTGGAARWASAKLERSHYLMHNYPAFMMEMKHVFPEDPORRE-----165
Qy 193 FVPKTFVHRIKE-----TDHQAIRKI-----MNOAYHKVMVHTTKELSKKXME 236
Db 166 -VAKRKIRLRQGMGVIDYSNAFQMIADLDWNEPALIDQYHEGLSDHIQEELS--HLE 222
Qy 237 HYEKVSS 243
Db 223 VAKLSLA 229

RESULT 15
US-09-736-457-1816
; Sequence 1816, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1816
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-736-457-1816

Query Match
Best Local Similarity 19.8%; Pred.No. 0.34; 91; Indels 62; Gaps 9;
Matches 49; Conservative 45; Mismatches 91; Indels 62; Gaps 9;
Qy 31 QOHNNTGESVLFHYPYKQPKNSHLVLIIEPKIEINKVIPESYQKEFEKSLFLQLS 90
Db 11 ESINNLRKYM-----KQSENNNLQSQVKLTENTTLREQVEPTPEDE-----D 56
Qy 91 SFLEKRGYSVSQFKD---ASEIPQDIKEKALLVLRMDGNVAILEDIVESDALSEKVID 147

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 16:33:52 ; Search time 30.6361 Seconds
(without alignments)
1511.612 Million cell updates/sec

Title: US-10-080-113-1

Perfect score: 1254

Sequence: 1 MKKGLAIVLGSLLASGAFY.....LSKKMEHYEKVSSEMKKKR 249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

-Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1254	100.0	249	10	US-09-881-752A-264
2	1254	100.0	249	12	US-10-080-113-1
3	201.5	16.1	260	10	US-09-396-973-4
4	199.5	15.9	260	10	US-09-396-973-2
5	199.5	15.9	260	10	US-09-848-967-1
6	132	10.5	207	10	US-09-881-752A-46
7	109.5	8.7	792	9	US-09-815-242-5724
8	108.5	8.7	1786	10	US-09-742-096-3
9	107.5	8.6	3878	12	US-10-080-608A-11
10	107.5	8.6	3899	15	US-10-171-311-4
11	107.5	8.6	3907	15	US-10-171-311-2
12	107.5	8.6	3911	12	US-10-370-685-100
13	107.5	8.6	3917	15	US-10-171-311-8
14	107.5	8.6	3925	15	US-10-171-311-6
15	106	8.5	648	9	US-09-815-242-10504

Sequence 8429, App
Sequence 1336, App
Sequence 168, App
Sequence 11057, A
Sequence 13855, A
Sequence 2, Appli
Sequence 5470, App
Sequence 12179, A
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 234, App
Sequence 234, App
Sequence 144, App
Sequence 1816, App
Sequence 1816, App
Sequence 1816, App
Sequence 28, Appli
Sequence 1816, App
Sequence 236, App
Sequence 246, App
Sequence 10, Appli
Sequence 29, Appli
Sequence 11389, A
Sequence 112, App
Sequence 3429, App
Sequence 11489, A

8.1 102 16
8.1 101.5 17
7.8 98 18
7.8 98 19
7.7 806 20
7.7 860 21
7.7 610 22
7.7 660 23
7.7 664 24
7.5 809 25
7.5 809 26
7.5 809 27
7.5 809 28
7.5 693 29
7.5 693 30
7.5 693 31
7.5 693 32
7.4 325 33
7.4 325 34
7.4 325 35
7.4 325 36
7.4 325 37
7.4 342 38
7.4 342 39
7.4 456 40
7.4 708 41
7.3 357 42
7.3 775 43
7.3 819 44
7.3 856 45

US-10-128-714-8429
US-10-083-357-1336
US-09-745-763-168
US-09-815-242-11057
US-10-156-761-13855
US-10-138-195-2
US-09-815-242-5470
US-09-815-242-12179
US-09-747-521-2
US-10-106-014-2
US-10-105-695-2
US-10-105-694-2
US-09-988-626-234
US-09-988-687-234
US-09-988-686-234
US-09-906-179A-144
US-09-736-457-1816
US-09-902-941-1816
US-09-849-626-1816
US-10-113-873-1816
US-10-382-600-28
US-10-017-754-1816
US-09-833-790-236
US-10-177-293-246
US-09-486-734A-10
US-10-382-600-29
US-09-815-242-11389
US-09-882-227-112
US-10-128-714-3429
US-09-815-242-11489

ALIGNMENTS

RESULT 1

US-09-881-752A-264
; Sequence 264, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Omen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1el Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 264
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-881-752A-264

Query Match 100.0%; Score 1254; DB 10; Length 249;
Best Local Similarity 100.0%; Pred. No. 2.1e-105;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKGLAIVLGSLLASGAFYTTALADGMPAKQHNNTGESVELHFHYPIKQEPKNSHLV 60
Db 1 MKKGLAIVLGSLLASGAFYTTALADGMPAKQHNNTGESVELHFHYPIKQEPKNSHLV 60
QY 61 VLIEPKIINKVPIESYOKFEKSLFLQLSFLERKGVSVQFQDASEIPQDIKEKALLV 120
Db 61 VLIEPKIINKVPIESYOKFEKSLFLQLSFLERKGVSVQFQDASEIPQDIKEKALLV 120

QY 121 LRMDGNVAILEDIVEESDALSEKVIDMSSGYLNFVPEKSEDIHHSFGIDVSKIKAVI 180
 Db 121 LRMDGNVAILEDIVEESDALSEKVIDMSSGYLNFVPEKSEDIHHSFGIDVSKIKAVI 180
 QY 181 ERVELRRTNSGGFVPEKTFVHRKETDHDQAIRKIMNQAYHKVMVHITKELSKKGMHEHYEK 240
 Db 181 ERVELRRTNSGGFVPEKTFVHRKETDHDQAIRKIMNQAYHKVMVHITKELSKKGMHEHYEK 240
 QY 241 VSSEMKKXK 249
 Db 241 VSSEMKKXK 249

RESULT 2

US-10-080-113-1
 ; Sequence 1, Application US/10080113
 ; Publication No. US20030166027A1
 ; GENERAL INFORMATION:
 ; APPLICANT: SACHS, GEORGE
 ; APPLICANT: VOLAND, PETRA
 ; TITLE OF INVENTION: COMPOSITIONS, TEST KITS AND METHODS FOR DETECTING
 ; TITLE OF INVENTION: HELICOBACTER PYLORI
 ; FILE REFERENCE: 626 06 PA
 ; CURRENT APPLICATION NUMBER: US/10/080,113
 ; CURRENT FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 249
 ; TYPE: PRT
 ; ORGANISM: Helicobacter pylori
 ; US-10-080-113-1

Query Match 100.0%; Score 1254; DB 12; Length 249;
 Best Local Similarity 100.0%; Pred. No. 2.1e-105;
 Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKGSLAIVGLSLASGAFYALADGMPAKQOHNNTGESVELHFPYPIKGQPKNSHLV 60
 Db 1 MKGSLAIVGLSLASGAFYALADGMPAKQOHNNTGESVELHFPYPIKGQPKNSHLV 60
 QY 61 VLIPEKLEINKVIPESYQKEFEKSLFLQSLFLEKGYVSQPKDASEIPQDIEKALLV 120
 Db 61 VLIPEKLEINKVIPESYQKEFEKSLFLQSLFLEKGYVSQPKDASEIPQDIEKALLV 120
 QY 121 LRMDGNVAILEDIVEESDALSEKVIDMSSGYLNFVPEKSEDIHHSFGIDVSKIKAVI 180
 Db 121 LRMDGNVAILEDIVEESDALSEKVIDMSSGYLNFVPEKSEDIHHSFGIDVSKIKAVI 180
 QY 181 ERVELRRTNSGGFVPEKTFVHRKETDHDQAIRKIMNQAYHKVMVHITKELSKKGMHEHYEK 240
 Db 181 ERVELRRTNSGGFVPEKTFVHRKETDHDQAIRKIMNQAYHKVMVHITKELSKKGMHEHYEK 240
 QY 241 VSSEMKKXK 249
 Db 241 VSSEMKKXK 249

RESULT 3

US-09-396-975-4
 ; Sequence 4, Application US/09396975
 ; Patent No. US20020168726A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bolin, Ingrid
 ; APPLICANT: Svennerholm, Ann-Mari
 ; TITLE OF INVENTION: Bacterial Antigens and Vaccine
 ; TITLE OF INVENTION: Compositions
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: White & Case
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York

; COUNTRY: U.S.A.
 ; ZIP: 10036-2787
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/396,975
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/669,560
 ; FILING DATE: 12-JUL-1996
 ; PRIOR APPLICATION DATA: PCT/SE96/00727
 ; APPLICATION NUMBER: 03-JUN-1996
 ; FILING DATE: 03-JUN-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: SE 9502007-9
 ; FILING DATE: 01-JUN-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: SE 9601085-5
 ; FILING DATE: 21-MAR-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Sterner, Richard J.
 ; REGISTRATION NUMBER: 35,372
 ; REFERENCE/DOCKET NUMBER: 1103326-0220
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 819-8200
 ; TELEFAX: (212)354-8113
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 260 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-396-975-4

Query Match 16.1%; Score 201.5; DB 10; Length 260;
 Best Local Similarity 24.8%; Pred. No. 3.4e-10;
 Matches 57; Conservative 66; Mismatches 78; Indels 29; Gaps 10;
 QY 39 SVELHFHPIKGPBKNSHLVLIPEKIEINKVIPESYQKEFEKSLFLQSLFLEKGY 98
 Db 39 ALKLNTH-PASEKVOALDEK-ILLRPAFYSDNIAKEYENKFKQIALKVEQILNQGY 96
 QY 99 SV---SQFKDASEIPQDIEKALILVRMDGNVAILED---IVEESD-----ALSEE 143
 Db 97 KVISVDSDDKDDFSPAQ--KKEGYLAVAMNGEIVLRPDPKETIOKKEPGLLFTGLDKM 154
 QY 144 KVIDMSSGYLNFVPEKSEDIHHSFGIDVSKIKAVIERVELRRT---NSGGFVPEKTFVH 200
 Db 155 EGVLIAPAGFIKVTILEPMSGESLDSFTMDLSELD--IQEKFYLTHTSHSGGLVST--- 208
 QY 201 RIKETDHDQAIRKIMNQAYHKVMVHITKELSKKGMHEHYEKVSSEMKKXK 249
 Db 209 MVKGTDSNDALKALNKIFANINQEIQDKLTKQNLKESYQKDAKELKGR 258

RESULT 4

US-09-396-975-2
 ; Sequence 2, Application US/09396975
 ; Patent No. US20020168726A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bolin, Ingrid
 ; APPLICANT: Svennerholm, Ann-Mari
 ; TITLE OF INVENTION: Bacterial Antigens and Vaccine
 ; TITLE OF INVENTION: Compositions
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: White & Case
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York

```

; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/396,975
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/669,560
; FILING DATE: 12-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/SE96/00727
; FILING DATE: 03-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9502007-9
; FILING DATE: 01-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9601085-5
; FILING DATE: 21-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Steiner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-0220
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8200
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 260 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-396-975-2

Query Match 15.9%; Score 199.5; DB 10; Length 260;
Best Local Similarity 24.8%; Pred. No. 5.1e-10;
Matches 57; Conservative 65; Mismatches 79; Indels 29; Gaps 10;

QY 39 SVELHYPYIKGQEPKNSHLVLIIEPKIEINKVIPESYQKEPKSLFLQLSSFLERKGY 98
Db 39 ALKLVNH-PASEKVALDEK-ILLRPAFOYSDNIAYENKFNQTLKVEQLQNGY 96
QY 99 SV-----SQPKDASEIPQDIKEKALLVLRMDGNVAILED-----IVEED-----ALSEE 143
Db 97 KVINVDSDKDDPSFAQ--KKEGYLAVANNGEIVLRDPDKRTIQKSEPGCLLFSTGLDGM 154
QY 144 KVIDMSSGYLNFVEPKSEDIHSGFDIVSKIKAVIERVELRRT---NSGGFVPKTFVH 200
Db 155 EGVLPAGFKVITILEPMSGESLDSFTMDLSELD--IQEKFLKTHSSHSGGLVST--- 208
QY 201 RIKETHDQAIRKIMNQAYHKVWHITKELSKKHMEHYKVSSEMKKRK 249
Db 209 MVKGTDSNDAIKSALNKIFGIMQEIIDKLTQKNLESYQKDAKELGKR 258

RESULT 5
US-09-848-967-1
; Sequence 1, Application US/09848967
; Publication No. US20020193295A1
; GENERAL INFORMATION:
; APPLICANT: CALENOFF, EMANUEL
; APPLICANT: DITLOW, CHARLES C.
; TITLE OF INVENTION: IMMUNOGENIC PEPTIDES AND USES THEREOF
; FILE REFERENCE: 2147/92141
; CURRENT APPLICATION NUMBER: US/09/848,967
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: 09/837,630
; PRIOR FILING DATE: 2001-04-18
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent In Ver. 2.1

```

```

; SEQ ID NO 1
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-848-967-1

Query Match 15.9%; Score 199.5; DB 10; Length 260;
Best Local Similarity 24.8%; Pred. No. 5.1e-10;
Matches 57; Conservative 65; Mismatches 79; Indels 29; Gaps 10;

QY 39 SVELHYPYIKGQEPKNSHLVLIIEPKIEINKVIPESYQKEPKSLFLQLSSFLERKGY 98
Db 39 ALKLVNH-PASEKVALDEK-ILLRPAFOYSDNIAYENKFNQTLKVEQLQNGY 96
QY 99 SV-----SQPKDASEIPQDIKEKALLVLRMDGNVAILED-----IVEED-----ALSEE 143
Db 97 KVINVDSDKDDPSFAQ--KKEGYLAVANNGEIVLRDPDKRTIQKSEPGCLLFSTGLDGM 154
QY 144 KVIDMSSGYLNFVEPKSEDIHSGFDIVSKIKAVIERVELRRT---NSGGFVPKTFVH 200
Db 155 EGVLPAGFKVITILEPMSGESLDSFTMDLSELD--IQEKFLKTHSSHSGGLVST--- 208
QY 201 RIKETHDQAIRKIMNQAYHKVWHITKELSKKHMEHYKVSSEMKKRK 249
Db 209 MVKGTDSNDAIKSALNKIFGIMQEIIDKLTQKNLESYQKDAKELGKR 258

RESULT 6
US-09-881-752A-46
; Sequence 46, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Comen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1el Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 207
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-881-752A-46

Query Match 10.5%; Score 132; DB 10; Length 207;
Best Local Similarity 25.9%; Pred. No. 0.00047;
Matches 55; Conservative 35; Mismatches 64; Indels 58; Gaps 11;

QY 42 LHFHYPIKGQEPKNSHLVLIIEPKIEIN-KVIPESYQKEPKSLFLQLSSFLERKGY 97
Db 24 LHLKY-----KDYPKNSALKTAFTLTTPKIFPNARFVPPFYQKEPKKAITQIAVFL---- 75
QY 98 YSVSQPKDASEIPQDIKEKALLVLRMDGNVAILEDIVEESDALSE--EKVIDMSSGYLN 154
Db 76 -----KKSAPILNVSGNVFFSFENPKDKLKAKEKLTIEFNA----- 115
QY 155 LNFVEPKSEDIHSGFDIVSKIKAVIER-----VELRRTNSGGFVPKTFV 199
Db 116 ----DPKA--VMRFLNLQASLILECVQTCTPDTLLIPTAFSPVYYANRLGDNPSLFS 169
QY 200 HRIKETHDQAIRKIMNQAYHKVWHITKELS 231
Db 170 QEDK-TYHN-ALIKALNKAYISLMGLEKRLN 199

```

```

RESULT 7
; Sequence 5724, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5724
; LENGTH: 792
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; NAME/KEY: VARIANT
; LOCATION: (1)...(792)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-5724

Query Match      8.7%; Score 109.5; DB 9; Length 792;
Best Local Similarity 22.3%; Pred. No. 0.34; Mismatches 90; Indels 79; Gaps 13;
Matches 61; Conservative 44;

Qy 16 SGAFYTLADGMPAKQOHNNTGESVLEHFPYIKGQEPKNS-----HLVVLIPEKIE 68
Db 536 SSCIIYSLDDVLRWSDNLAENQL---YSLDAQRLSQSSSLFNKRVKQIVVKAQRIS 592
Qy 69 INKVIPESYQKEPKSLFLQSSFLERKGYVSQFKDASEIPQDIKEKALLVLRDGNVA 128
Db 593 EQGVKAREMANEPKSLISQRLDLYVEERN-RVLEIDDAEN--RDFKALAKDVFEK 644
Qy 129 ILEDIVEESDALSEEKVIDMSSGYLNLNFVEPKSEDIHSFGIDVSKI-----KAVI 180
Db 645 ----FVNEEKVLTKSRVVEI--YQNLSP-----QFNKDVACVNFKQAVVFTL 688
Qy 181 -----ERVELRTNSGG-----FVPTKTFVHRI-----KETDHDQAIRKIMO----- 217
Db 689 LEQPEKQVALNRKMQSAYTYNIFVQKVFLEKALDSCWLEQVDYLOQLKASVNRQGRN 748
Qy 218 ---AYHKVMVHITKELSKHMEHYEKVSSEMKKR 248
Db 749 AIEFYHRVA-----LDSFEVWTRNIKKR 771

RESULT 8
; Sequence 3, Application US/09742096
; Patent No. US20020155441A1
; GENERAL INFORMATION:
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 200773USODIV
; CURRENT APPLICATION NUMBER: US/09/742,096
; CURRENT FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 08/973,642
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: PCT/FR96/00894
; PRIOR FILING DATE: 1996-06-12
; PRIOR APPLICATION NUMBER: FR 95/07007
; PRIOR FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: P. falciparum
US-09-742-096-3

Query Match      8.7%; Score 108.5; DB 10; Length 1786;
Best Local Similarity 26.0%; Pred. No. 1.4;
Matches 50; Conservative 35; Mismatches 90; Indels 17; Gaps 7;

Qy 62 LIEPKIEINKVIPESYQKEPKSLFLQSSFLERKGYVSQFKDASEIPQDIKEKALLV 121
Db 1355 ILEEKKEIEKHDPKFEKEEAEIKDLADILKEYSSLEVEEKKLEEV-HELKEEVEHII 1413
Qy 122 RMDGNV-AILEDIVEESDALSEEKVIDMSSGYLNLNFVEPKS-EDIHSFGIDVSKIKAV 179
Db 1414 SCDAHIKGLEDDLEEVDDL-KGSILDMLKGMELGMDKESLEDVTTKLGERVESELDKV 1472
Qy 180 IERV-----ELRRTNSGGFVPTKTFVHRIKETDHDQAIRKINQAYHKVMVHITKELSKKH 234
Db 1473 LSSALGMDEEQMTRKKAQRPKEVLLKEEVEKKKTKK---KVRFDIKDKPKD- 1528
Qy 235 MEHYEKVSSEM 246
Db 1529 ----EIVEEMK 1536

RESULT 9
US-10-080-608A-11
; Sequence 11, Application US/10080608A
; Publication No. US20030198956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/10/080,608A
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-080-608A-11

Query Match      8.6%; Score 107.5; DB 12; Length 3878;
Best Local Similarity 20.7%; Pred. No. 5;
Matches 52; Conservative 47; Mismatches 101; Indels 51; Gaps 10;

Qy 14 LASGAFYTLADGMPAKQOHNNTGESVLEHFPYIKGQEPKNSHLVVLIPEKIEINKVI 73
Db 263 LRNSTHSSTAADLLQAKQOILTHOQQLS-----EQDHL----- 296
Qy 74 PESYQKEPKSLFLQSSFLER--KGYVSQFKDA-SEIPQDIKEKALLVLRMDGNVAIL 130

```

```

Db 297 -EDYQK--KKEDFTMOISFLQEKIKYVEEQDKVNSKKEBIOQKETTIEEL--NTKII 351
QY 131 ED---IVESDALS-----BEKVIDMSSGYLNLFVPEKSEDIHSPGIDVSKIKA 178
Db 352 EEEKKTLEKDKLTADKLLGELQEQIVQKNQEIKNMKLELTSKQKQSQSEEEKQLMG 411
QY 179 VIERVELRRTNSGGFVPEKTFVHRIKETDHD-QAIRKIMNOAYHKVWVHTKLSKKMEH 237
Db 412 TVEELQKRNHKDSQSFETDIVORMEQETQKLEQLRAELDEMYGQOIVQMKQLIRQHMAQ 471
QY 238 YEKVSSEMKKR 248
Db 472 ME----EMKTR 478

```

```

RESULT 10
US-10-171-311-4
; Sequence 4, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 3899
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-4

```

```

Query Match 8.6%; Score 107.5; DB 15; Length 3899;
Best Local Similarity 20.7%; Pred. No. 5.1; Indels 51; Gaps 10;
Matches 52; Conservative 47; Mismatches 101;

QY 14 LASGAFYTTALADGMPAKQOHNNTGESVELHFPYKQEPKNSHLVLLIEPKIEINKVI 73
Db 251 LRNSTHSSTAADLLQAKQILTHQOQLE-----EQDHLL----- 284
QY 74 PESYCKEPEKSLFLQSSFLER--KGYSVQPKDA-SEIPQDIKEKALLVLRMDGNVAIL 130
Db 285 -EDYQK--KKEDFTMOISFLQEKIKYVEEQDKVNSKKEBIOQKETTIEEL--NTKII 339
QY 131 ED---IVESDALS-----BEKVIDMSSGYLNLFVPEKSEDIHSPGIDVSKIKA 178
Db 340 EEEKKTLEKDKLTADKLLGELQEQIVQKNQEIKNMKLELTSKQKQSQSEEEKQLMG 399
QY 179 VIERVELRRTNSGGFVPEKTFVHRIKETDHD-QAIRKIMNOAYHKVWVHTKLSKKMEH 237
Db 400 TVEELQKRNHKDSQSFETDIVORMEQETQKLEQLRAELDEMYGQOIVQMKQLIRQHMAQ 459
QY 238 YEKVSSEMKKR 248
Db 460 ME----EMKTR 466

```

```

RESULT 11
US-10-171-311-2
; Sequence 2, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 3907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-2

```

```

Query Match 8.6%; Score 107.5; DB 15; Length 3907;
Best Local Similarity 20.7%; Pred. No. 5.1; Indels 51; Gaps 10;
Matches 52; Conservative 47; Mismatches 101;

QY 14 LASGAFYTTALADGMPAKQOHNNTGESVELHFPYKQEPKNSHLVLLIEPKIEINKVI 73
Db 251 LRNSTHSSTAADLLQAKQILTHQOQLE-----EQDHLL----- 284
QY 74 PESYCKEPEKSLFLQSSFLER--KGYSVQPKDA-SEIPQDIKEKALLVLRMDGNVAIL 130
Db 285 -EDYQK--KKEDFTMOISFLQEKIKYVEEQDKVNSKKEBIOQKETTIEEL--NTKII 339
QY 131 ED---IVESDALS-----BEKVIDMSSGYLNLFVPEKSEDIHSPGIDVSKIKA 178
Db 340 EEEKKTLEKDKLTADKLLGELQEQIVQKNQEIKNMKLELTSKQKQSQSEEEKQLMG 399
QY 179 VIERVELRRTNSGGFVPEKTFVHRIKETDHD-QAIRKIMNOAYHKVWVHTKLSKKMEH 237
Db 400 TVEELQKRNHKDSQSFETDIVORMEQETQKLEQLRAELDEMYGQOIVQMKQLIRQHMAQ 459
QY 238 YEKVSSEMKKR 248
Db 460 ME----EMKTR 466

```

```

RESULT 12
US-10-370-685-100
; Sequence 100, Application US/10370685
; Publication No. US20030215903A1
; GENERAL INFORMATION:
; APPLICANT: Hyman, Paul
; APPLICANT: Goldberg, Edward
; TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements
; FILE REFERENCE: NANF.P-004
; CURRENT APPLICATION NUMBER: US/10/370,685
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: 10/080,608
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PatentIn version 3.2

```

```
; SEQ ID NO 100
; LENGTH: 3911
; TYPE: PRT
; ORGANISM: human
US-10-370-685-100

Query Match      8.6%; Score 107.5; DB 12; Length 3911;
Best Local Similarity 20.7%; Pred. No. 5.1;
Matches 52; Conservative 47; Mismatches 101; Indels 51; Gaps 10;

QY 14 LASGAFY TALADGMPAKOQHNNNTGESVELHFPYPIKQBPKNHVLVLLIEPKIEINKVI 73
DB 263 LNSTHSTAADLLQAKQILTHQQOLE-----EQDHL----- 296

QY 74 PESYQKEFEKSLFLQLSSFLER--KGYSVSQFKDA--SEIPQDIKEKALLVLRMDGNVAIL 130
DB 297 -EDYQK--KKEDFTWQISFLQEKIKVYEMEQDKKVENSKKEEIOEKETIIEEL--NTKII 351

QY 131 ED--IIVESDAL--BEKVIDMSSGYLNLFNVEPKSEDIHHSFGIDVSKIKA 178
DB 352 EEKKTLELKDKLTTADKLLGELQEQIVQKNQEIKNMKLELTNSKQKERSSEIKQLMG 411

QY 179 VIERVELRRTNSGGFVPKTFVHRIKETDHD--QAIRKIMNQAYHKVMVHITKELSKHMEH 237
DB 412 TVEELQKRNHKDSQFETDIVORMEQETQKLEQLRAELDEMYGQOIVQMKEQLIRQHMAQ 471

QY 238 YEKVSSEMKR 248
DB 472 ME----EMKTR 478

RESULT 13
US-10-171-311-8
; Sequence 8, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Marjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; PRIOR FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-6

Query Match      8.6%; Score 107.5; DB 15; Length 3925;
Best Local Similarity 20.7%; Pred. No. 5.1;
Matches 52; Conservative 47; Mismatches 101; Indels 51; Gaps 10;

QY 14 LASGAFY TALADGMPAKOQHNNNTGESVELHFPYPIKQBPKNHVLVLLIEPKIEINKVI 73
DB 251 LNSTHSTAADLLQAKQILTHQQOLE-----EQDHL----- 284

QY 74 PESYQKEFEKSLFLQLSSFLER--KGYSVSQFKDA--SEIPQDIKEKALLVLRMDGNVAIL 130
DB 285 -EDYQK--KKEDFTWQISFLQEKIKVYEMEQDKKVENSKKEEIOEKETIIEEL--NTKII 339

QY 131 ED--IIVESDAL--BEKVIDMSSGYLNLFNVEPKSEDIHHSFGIDVSKIKA 178
DB 340 EEKKTLELKDKLTTADKLLGELQEQIVQKNQEIKNMKLELTNSKQKERSSEIKQLMG 399

QY 179 VIERVELRRTNSGGFVPKTFVHRIKETDHD--QAIRKIMNQAYHKVMVHITKELSKHMEH 237
DB 400 TVEELQKRNHKDSQFETDIVORMEQETQKLEQLRAELDEMYGQOIVQMKEQLIRQHMAQ 459

QY 238 YEKVSSEMKR 248
DB 460 ME----EMKTR 466

US-10-171-311-6
; Sequence 6, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Marjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; PRIOR FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-6

Query Match      8.6%; Score 107.5; DB 15; Length 3925;
Best Local Similarity 20.7%; Pred. No. 5.1;
Matches 52; Conservative 47; Mismatches 101; Indels 51; Gaps 10;

QY 14 LASGAFY TALADGMPAKOQHNNNTGESVELHFPYPIKQBPKNHVLVLLIEPKIEINKVI 73
DB 251 LNSTHSTAADLLQAKQILTHQQOLE-----EQDHL----- 284

QY 74 PESYQKEFEKSLFLQLSSFLER--KGYSVSQFKDA--SEIPQDIKEKALLVLRMDGNVAIL 130
DB 285 -EDYQK--KKEDFTWQISFLQEKIKVYEMEQDKKVENSKKEEIOEKETIIEEL--NTKII 339

QY 131 ED--IIVESDAL--BEKVIDMSSGYLNLFNVEPKSEDIHHSFGIDVSKIKA 178
DB 340 EEKKTLELKDKLTTADKLLGELQEQIVQKNQEIKNMKLELTNSKQKERSSEIKQLMG 399

QY 179 VIERVELRRTNSGGFVPKTFVHRIKETDHD--QAIRKIMNQAYHKVMVHITKELSKHMEH 237
DB 400 TVEELQKRNHKDSQFETDIVORMEQETQKLEQLRAELDEMYGQOIVQMKEQLIRQHMAQ 459

QY 238 YEKVSSEMKR 248
DB 460 ME----EMKTR 466
```

RESULT 15

US-09-815-242-10504
; Sequence 10504, Application US/09815242
; Patent NO. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Oalsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10504
; LENGTH: 648
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10504

Query Match 8.5%; Score 106; DB 9; Length 648;
Best Local Similarity 22.3%; Pred.No. 0.54; Indels 60; Gaps 9;
Matches 42; Conservative 32; Mismatches 54
QY 103 FKDASEIPQDIKEKALLVLRMDGNVAILEDIVEESDALSEKVIDMSSGYNLNFVEPKS 162
DB 91 FADILEMEQRMRELETKISEMDPTTSVYEGILKEYDQL--QHTFSEKGY----GVENEI 144
QY 163 EDIIHSGIDVS-----KIKAVIERVELRR-----TNS-----G 191
DB 145 RSVLHGFGFDESPTTKDIOTLSGGQKTRLALARMLOKRPDILIDEPINHLDIETLSWLE 204
QY 192 GFVPKTFVHRIKETDHDQ-AIRKIMNQAYHKVMVHITKELSKGMEHY----- 238
DB 205 SYLP-SVAGALLIVSHDRYFLDKVNEVY-----ELSRKQWTHYKGNYSKYLELKA 254
QY 239 EKVSEMK 246
DB 255 EQLASEWK 262

Search completed: December 10, 2003, 18:40:30
Job time : 38.6361 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:30:22 ; Search time 16.6217 Seconds
(without alignments)
1440.645 Million cell updates/sec

Title: US-10-080-113-1
Perfect score: 1254
Sequence: 1 MKKGSIAVLGSLASGAFY.....LSKKMEHYKVSSEMKKKK 249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR 76:*
1: Piri:*
2: Piri2:*
3: Piri3:*
4: Piri4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1254	100.0	249	B64571	probable neuraminyl
2	1199	95.6	249	E71865	probable neuraminyl
3	302.5	24.1	278	D64581	hypothetical prote
4	300	23.9	282	A71930	probable neuraminyl
5	198.5	15.8	260	E64619	adhesin A - Helico
6	194.5	15.5	260	C71896	neuraminylactose-
7	132	10.5	207	A64655	hypothetical prote
8	129	10.3	207	E71943	hypothetical prote
9	115.5	9.2	722	AD1796	hypothetical membr
10	108.5	8.7	1558	E71503	RESA-H3 antigen PF
11	108	8.6	1642	C70416	pd377 protein - En
12	107.5	8.6	1642	T08880	NMDA receptor-bind
13	106	8.5	340	F89852	extracellular ECM
14	104.5	8.3	796	A90073	hypothetical prote
15	103.5	8.3	864	B90395	purine NTPase (imp
16	102	8.1	1516	E71619	PAD2 endonuclease
17	101.5	8.1	312	S66952	hypothetical prote
18	101.5	8.1	478	C70416	trigger factor tlg
19	100.5	8.0	576	S27750	SLP1 protein homol
20	100	8.0	880	F75103	conserved hypothet
21	99	7.9	670	E70145	periplasmic protei
22	99	7.9	762	AF1610	DNA translocase ho
23	98.5	7.9	757	AB1248	DNA translocase ho
24	98.5	7.9	3770	A40889	delta-(L-alpha-ami
25	98	7.8	806	A40992	ferritin precursor
26	98	7.8	806	A64078	DNA topoisomerase
27	98	7.8	840	H71114	probable cell divi
28	98	7.8	903	C64444	cell division cont
29	97.5	7.8	403	AI1636	thiamin biosynthes

ftsH proteinase ac
thiamin biosynthes
transcription-repa
araC-type sugar me
B447 protein - pha
endopeptidase Clp
glutinin - human
glutinin - human
recombination prot
hypothetical prote
conserved hypothet
anthrax toxin leth
type I restriction
ig V-region-like B
hypothetical prote
hypothetical prote

ALIGNMENTS

RESULT 1

B64571

probable neuraminylactose-binding hemagglutinin jhp0971 - Helicobacter pylori (strain 261

C:Species: Helicobacter pylori

C>Date: 09-Aug-1997 #sequence_revision 15-Aug-1997 #text_change 08-Oct-1999

C:Accession: B64571

R: Tomb, J.F.; White, O.; Kervilave, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Katpik, P.D.; Smith, H.O.; Fraser, C.

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: B64571

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-249 <TOM>

A:Cross-references: GB:AE000557; GB:AE000511; NID:G2313514; PIDN:AD07478.1; PID:G231351

Query Match 100.0%; Score 1254; DB 2; Length 249;
Best Local Similarity 100.0%; Pred. No. 5.8e-76;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MKKGSIAVLGSLASGAFY	TALADGMPAKOCHNNTGESVELHFPYIKGQEPKNSHLV	60
DB	1	MKKGSIAVLGSLASGAFY	TALADGMPAKOCHNNTGESVELHFPYIKGQEPKNSHLV	60
QY	61	VLIPEKIEINKVIPESYQKEPKSLFLQLSSFLERKGYSVSQFKDASEIPODIKEKALLV	120	
DB	61	VLIPEKIEINKVIPESYQKEPKSLFLQLSSFLERKGYSVSQFKDASEIPODIKEKALLV	120	
QY	121	LRMDGNVALLEDIVESDSEKVIDMSSGYLNINLNFVEPKSEDIHISFGIDVSKIKAVI	180	
DB	121	LRMDGNVALLEDIVESDSEKVIDMSSGYLNINLNFVEPKSEDIHISFGIDVSKIKAVI	180	
QY	181	ERVELRRTNSGGVPKTFVHRIKETDHDQAIKRMNQAYHKVMVHITKELSKGMEHYEK	240	
DB	181	ERVELRRTNSGGVPKTFVHRIKETDHDQAIKRMNQAYHKVMVHITKELSKGMEHYEK	240	
QY	241	VSEMKKKK	249	
DB	241	VSEMKKKK	249	

RESULT 2

E71865

probable neuraminylactose-binding hemagglutinin jhp0971 - Helicobacter pylori (strain J

C:Species: Helicobacter pylori

A:Variety: strain J99

C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C:Accession: E71865

R: Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.,

Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: E71865
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-249 <ARN>
A:CROSS-references: GB:AE001526; GB:AE001439; NID:94155550; PIDN:AA06549.1; PID:9415555
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0971

Query Match 95.6%; Score 1199; DB 2; Length 249;
Best Local Similarity 94.8%; Pred. No. 2.5e-72;
Matches 236; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 1 MKKSLAIVLGSLLASGAFYTLADGMPAKQOHNNTGESVELHFHYPIKQKPPKSHLV 60

DB 1 MKKSLAIVLGSLLASGAFYTLADGMPAKQOHNNTGESVELHFHYPIKQKPPKSHLV 60

QY 61 VLIPEKIEINKVIPESYQKEPEKSLFLQSSFLERKGYVSQKDAISEIPQDIKEKALLV 120

DB 61 VLIPEKIEINKVIPESYQKEPEKSLFLQSSFLERKGYVSQKDAISEIPQDIKEKALLV 120

QY 121 LRMDGNVAILEDIVVEESDALSEEKVIDMSSGYLNLNFEVPEKSEDIHSPGIDVSKIKAVI 180

DB 121 LRMDGNVAILEDIVVEESDALSEEKVIDMSSGYLNLNFEVPEKSEDIHSPGIDVSKIKAVI 180

QY 181 ERVELRRTNSGGFVPKTFVHRKETDHDQAIKIMQAYHKVMVHITKELSKGMEHYEK 240

DB 181 ERVELRRTNSGGFVPKTFVHRKETDHDQAIKIMQAYHKVMVHITKELSKGMEHYEK 240

QY 241 VSSEMKKRRK 249

DB 241 VSSEMKKRRK 249

RESULT 3
D64581
Hypothetical protein HP0492 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: D64581
R:Tomb, J.F.; White, O.; Kervilave, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Lofthus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: D64581
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-278 <TOM>
A:CROSS-references: GB:AE000564; GB:AE000511; NID:92313602; PIDN:AA07568.1; PID:9231361

Query Match 24.1%; Score 302.5; DB 2; Length 278;
Best Local Similarity 29.1%; Pred. No. 5.1e-13;
Matches 73; Conservative 57; Mismatches 110; Indels 11; Gaps 5;

QY 4 GSLAIVLGSLLASGAFYTLADGMPAKQOHNNTGESVE---LHFHYPIKQKPPKSHLV 59

DB 21 GLSSVLIGCANPSPAEKTPNDKQVQVOTHERMKTSEHVTPLDFNYPHIVQAPQNHV 80

QY 60 VLIPEKIEINKVIPESYQKEPEKSLFLQSSFLERKGYVSQKDAISEIPQDIKEKALL 119

DB 81 VGILTPRIQVSDNL-KPYIDKFQDALINQITIFEKRGYQVLRFDQEKALNAQDKRIFS 139

QY 120 VLRMDGNVAILEDI---VEESDALSEEKVIDMSSGYLNLNFEVPEKSEDIHSPGIDVSKI 176

DB 140 VLDLKGWVGILEDLKMLKDPNPNLTLVDQSSGSGWFWNFYEPESNRVVDHFAVEVGTTF 199

QY 177 KAVIERVELRRTNSGGF-VPKTFVHRKETDHDQAIKIMQAYHKVMVHITKELSKGHM 235
DB 200 QAM---TYYTKNNSSGLNSSLIIHEYLEKKNEDAIHKILNRVAVVMKAVTELTKENI 257
QY 236 EHYEKVSSEMK 246
DB 258 DKYREAIRDMRK 268
RESULT 4
A71930
Probable neuraminylactose-binding hemagglutinin jhp0444 - Helicobacter pylori (strain J
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: A71930
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, B.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: A71930
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-282 <ARN>
A:CROSS-references: GB:AE001479; GB:AE001439; NID:94154979; PIDN:AA06033.1; PID:9415499
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0444

Query Match 23.9%; Score 300; DB 2; Length 282;
Best Local Similarity 29.7%; Pred. No. 7.7e-13;
Matches 76; Conservative 57; Mismatches 107; Indels 16; Gaps 6;

QY 6 LAIVLGSLLASGAFYTLADGMP--AKQO-----HNNTGESVE---LHFHYPIKQKQEP 54
DB 18 VALGSSVLIGCANPSPAEKTPNDKQVQVOTHERMKTSEHVTPLDFNYPHIVQAP 77
QY 55 KNSHLVLIPEKIEINKVIPESYQKEPEKSLFLQSSFLERKGYVSQKDAISEIPQDIK 114
DB 78 QNHVVGILMPRIQVSDNL-KPYIDKFQDALINQITIFEKRGYQVLRFDQEKALNVQDK 136
QY 115 EKALLVLRMDGNVAILEDI---VEESDALSEEKVIDMSSGYLNLNFEVPEKSEDIHSPGFI 171
DB 137 KPIFVLDKMGVGLLEDLKMLKDPNPNLTLVDQSSGSGWFWNFYEPESNRVVDHFAV 196
QY 172 DYSKIKAVIERVELRRTNSGGF-VPKTFVHRKETDHDQAIKIMQAYHKVMVHITKEL 230
DB 197 EVGTQAITVTVTSTNNASGSGSSKSVIHENLDKREDALHKILNRVAVVMKAVTEL 256
QY 231 SKKMEHYEKVSSEMK 246
DB 257 TKNIAKYRDAIRDMRK 272

RESULT 5
E64619
adhesin A - Helicobacter pylori (strains 26695 and 8826)
N:Alternate names: flagellar sheath adhesin hpa; N-acetylneuraminylactose-binding fibr
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: E64619; B47052; C47058
R:Tomb, J.F.; White, O.; Kervilave, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Lofthus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: E64619
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-260 <TOM>

A;Cross-references: GB:AE000591; GB:AE000511; NID:G2313918; PIDN:AAD07844.1; PID:G2313921
A;Experimental source: strain 26695
R;Evans, D.G.; Karjalainen, T.K.; Evans Jr., D.J.; Graham, D.Y.; Lee, C.H.
J. Bacteriol. 175, 674-683, 1993
A;Title: Cloning, nucleotide sequence, and expression of a gene encoding an adhesin subunit
A;Reference number: A47052; MUID:93139035; PMID:7678592
A;Accession: B47052
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-206
A;Cross-references: GB:X61574; NID:G732735
A;Note: sequence extracted from NCBI backbone (NCBI:123532, NCBIP:123534); sequence in
A;Accession: C47052
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-206
A;Cross-references: GB:X61574; NID:G732735
A;Note: sequence extracted from NCBI backbone (NCBI:123532, NCBIP:123535)
R;Evans, D.G.; Lampert, H.C.; Nakano, H.; Eaton, K.A.; Burnens, A.P.; Bronsdon, M.A.; Evans
Gene 163, 97-102, 1995
A;Title: Genetic evidence for host specificity in the adhesin-encoding genes hxaA of Helicobacter pylori
A;Reference number: PC4088; MUID:96001251; PMID:7557486
A;Accession: PC4088
A;Molecule type: DNA
A;Residues: 43-61, 'K', 63-83, 'T', 85-88, 'E', 90-99, 'N', 101-108, 'F', 110-111, 'A', 113-123, 'I',
A;Experimental source: strain 8926
C;Genetics:
F;134-139/Region: receptor binding #status predicted
Query Match 15.8%; Score 198.5; DB 2; Length 260;
Best Local Similarity 24.8%; Pred. No. 3.5e-06;
Matches 57; Conservative 65; Mismatches 79; Indels 29; Gaps 10;
QY 39 SVLHHPYIKGQEPKNSHLVLIPEKINVKIPESYQKEFKSLFLQSSFLERKGY 98
Db 39 ALKLNH-PASEKQVQALDEK-ILLRPAFYSDNIKEYNKQTKALKEVILQNGY 96
QY 99 SV-----SQFKDASEIPQDIKEKALLVLRMDGNVAILED-----IVESD-----ALSEE 143
Db 99 KVINVDSSDKDLSFQ--KKEGYLAVAMNGEIVLRPDPKRTIQKKEPGLFSTGLDKM 154
QY 144 KVIDMSSGYLNFVPEKEDIHSGIDVSKIKAVIERVELRRT---NSGGFVPKTFVH 200
Db 144 EGVLPAGFVKVITLFPMSGESLDSFTMDLSELD--IQKFLKTHSSHSGGLVST--- 208
QY 201 RIKETDH-DOAIRKIMNQAYHKVMVHITKELSKKMEHYKVSSEMKRK 249
Db 209 MVKGTNSDAIKSALNKIFANIMQEIQKLTQKNLESYQKDAKELGKR 258
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: A64520
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-207 <TOM>
A;Cross-references: GB:AE000615; GB:AE000511; NID:G2314230; PIDN:AAD08133.1; PID:G2314231

Query Match 10.5%; Score 132; DB 2; Length 207;
Best Local Similarity 25.9%; Pred. No. 0.066;
Matches 55; Conservative 35; Mismatches 64; Indels 58; Gaps 11;
QY 42 LHPHYPIKGQEPKNSHLVLIPEKIN-KVIPESYQKEFKSLFLQSSFLERK 97
Db 24 LHLK-----KDYKNSALKTAFTLTPKIFPNARFVFFYQKEFKKAITQIAYFL----- 75
QY 98 YSVSQFKDASEIPQDIKEKALLVLRMDGNVAILEDIVESDSEALSE--EKVIDMSSGYLN 154
Db 98 -----KDKSAFILNVSNGVFFSFEENPKDLKAIKERLKTIEPNA---- 115
QY 155 LNFVEPKSEDIHSGIDVSKIKAVIER-----VELRRTNSGGFVPKTFV 199
Db 116 ----DPKA--VMRFNLQASLLILECVPTTCPPDTLLIPTAFVVPVYANRLGDNPSLFS 169
QY 200 HRIKETDHDQAIRKIMNQAYHKVMVHITKEL 231
Db 170 QEDK-TYHN-ALIKALNKAYVSLMEGLEKRLN 199
RESULT 8
B71943
Hypothetical protein jhp0344 - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
A;Accession: B71943
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: B71943

Query Match 15.5%; Score 194.5; DB 2; Length 260;
Best Local Similarity 24.8%; Pred. No. 3.5e-06;
Matches 57; Conservative 65; Mismatches 79; Indels 29; Gaps 10;
QY 39 SVLHHPYIKGQEPKNSHLVLIPEKINVKIPESYQKEFKSLFLQSSFLERKGY 98
Db 39 ALKLNH-PASEKQVQALDEK-ILLRPAFYSDNIKEYNKQTKALKEVILQNGY 96
QY 99 SV-----SQFKDASEIPQDIKEKALLVLRMDGNVAILED-----IVESD-----ALSEE 143
Db 99 KVINVDSSDKDLSFQ--KKEGYLAVAMNGEIVLRPDPKRTIQKKEPGLFSTGLDKM 154
QY 144 KVIDMSSGYLNFVPEKEDIHSGIDVSKIKAVIERVELRRT---NSGGFVPKTFVH 200
Db 144 EGVLPAGFVKVITLFPMSGESLDSFTMDLSELD--IQKFLKTHSSHSGGLVST--- 208
QY 201 RIKETDH-DOAIRKIMNQAYHKVMVHITKELSKKMEHYKVSSEMKRK 249
Db 209 MVKGTNSDAIKSALNKIFANIMQEIQKLTQKNLESYQKDAKELGKR 258
RESULT 6
C71896
neuraminylactose-binding hemagglutinin precursor - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
A;Accession: C71896
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: C71896
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-260 <ARN>
A;Cross-references: GB:AE001504; GB:AE001439; NID:G4155285; PIDN:AAD06306.1; PID:G415528
A;Experimental source: strain J99
C;Genetics:
A;Gene: hpaA

Best Local Similarity 23.9%; Pred. No. 6.5e-06;
Matches 55; Conservative 67; Mismatches 79; Indels 29; Gaps 10;
QY 39 SVLHHPYIKGQEPKNSHLVLIPEKINVKIPESYQKEFKSLFLQSSFLERKGY 98
Db 39 ALKLNH-PASEKQVQALDEK-ILLRPAFYSDNIKEYNKQTKALKEVILQNGY 96
QY 99 SV-----SQFKDASEIPQDIKEKALLVLRMDGNVAILED-----IVESD-----ALSEE 143
Db 97 KVINVDSSDKDLSFQ--KKEGYLAVAMNGEIVLRPDPKRTIQKKEPGLFSTGLDKM 154
QY 144 KVIDMSSGYLNFVPEKEDIHSGIDVSKIKAVIERVELRRT---NSGGFVPKTFVH 200
Db 155 EGVLPAGFVKVITLFPMSGESLDSFTMDLSELD--IQKFLKTHSSHSGGLVST--- 208
QY 201 RIKETDH-DOAIRKIMNQAYHKVMVHITKELSKKMEHYKVSSEMKRK 249
Db 209 MVKGTNSDAIKSALNKIFANIMQEIQKLTQKNLESYQKDAKELGKR 258
RESULT 7
A64655
Hypothetical protein HP1081 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
A;Accession: A64655
R;Tomb, J.F.; White, O.; Kervatage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney,
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.;
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: A64520
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-207 <TOM>
A;Cross-references: GB:AE000615; GB:AE000511; NID:G2314230; PIDN:AAD08133.1; PID:G2314231

Query Match 10.5%; Score 132; DB 2; Length 207;
Best Local Similarity 25.9%; Pred. No. 0.066;
Matches 55; Conservative 35; Mismatches 64; Indels 58; Gaps 11;
QY 42 LHPHYPIKGQEPKNSHLVLIPEKIN-KVIPESYQKEFKSLFLQSSFLERK 97
Db 24 LHLK-----KDYKNSALKTAFTLTPKIFPNARFVFFYQKEFKKAITQIAYFL----- 75
QY 98 YSVSQFKDASEIPQDIKEKALLVLRMDGNVAILEDIVESDSEALSE--EKVIDMSSGYLN 154
Db 98 -----KDKSAFILNVSNGVFFSFEENPKDLKAIKERLKTIEPNA---- 115
QY 155 LNFVEPKSEDIHSGIDVSKIKAVIER-----VELRRTNSGGFVPKTFV 199
Db 116 ----DPKA--VMRFNLQASLLILECVPTTCPPDTLLIPTAFVVPVYANRLGDNPSLFS 169
QY 200 HRIKETDHDQAIRKIMNQAYHKVMVHITKEL 231
Db 170 QEDK-TYHN-ALIKALNKAYVSLMEGLEKRLN 199
RESULT 8
B71943
Hypothetical protein jhp0344 - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
A;Accession: B71943
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: B71943

A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-207 <ARN>
A:Cross-references: GB:AE001470; GB:AE001439; NID:94154869; PIDN:AND05923.1; PID:9415487
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0344

Query Match 10.3%; Score 129; DB 2; Length 207;
Best Local Similarity 25.0%; Pred. No. 0.1; Indels 58; Gaps 10;
Matches 53; Conservative 34; Mismatches 67; Indels 58; Gaps 10;

QY 42 LHFYPIKGQEPKNSHL---VWLEPKIEIN-KVPIESYQKFEKSLFLQLSSFLERKG 97

Db 24 LHLXY---KDYKNSPLKASTLTPPKIFPNAHFVPPFYQKFKALAQIAVFL--- 75

QY 98 YSVSQPKDASHPDQIEKALLVLRMDGNVAI-LEDIVESDALSE--EKVIDMSSGYLN 154

Db 76 -----KDKGALTNISGNVFFSFESPKDLKAIKRLKKTIEPNT----- 115

QY 155 LNFVEPKSEDIHSGIDVSKIAVIER-----VELRRTNSGGFVPKTFV 199

Db 116 ---DPKA--VMRELNLQASILECVQPTACPPDTLLIPTALSVPIDYANRLGDNPSLFP 169

QY 200 HRIKETHDQAIRKIMNQAHKVMVHTKELS 231

Db 170 Q--EDKSYHNALIKALNKAYSLMEGLEKRLN 199

RESULT 9

AD1796
hypochemical membrane protein lin2915 (imported) - Listeria innocua (strain Clip11262)

C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

C:Accession: AD1796
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Krett, J.; Kuhn, M.; Kunst, F.; Kurapk, G.; Madueno, E.; Maitournan, A.; Ma

ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AD1796

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-722 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC98140.1; PID:gl6415456; GSPDB:GN00178

A:Experimental source: strain Clip11262

C:Genetics:

A:Gene: lin2915

Query Match 9.2%; Score 115.5; DB 2; Length 722;

Best Local Similarity 28.4%; Pred. No. 3.9;

Matches 63; Conservative 35; Mismatches 75; Indels 49; Gaps 14;

QY 37 GESVELHFHY---PIKKGQEPKNSHLVLEPKIEIN-KVPIESYQKFEKSL-FLQLSS 91

Db 427 GNSVEVDNFYLNLPKPSAQ-GQVNEALINDKV-LNIIVPTS-KKEPKIKKAYLDH 483

QY 92 FLERKGVSVQFKDASHPDQIEKALLVLRMDG-----NVAILEDIVESDALSEEKVIDM 148

Db 484 FYQKVEVANNYNEALNRP-----VLESKDDLINIIYAQN---NQDYFS-----YDS 529

QY 149 SSGVLNL-NEVEPKSEDIHSGIDVSKIAVIERVELRRTNSGGFVPKTFVHRIKETDH 207

Db 530 SAGDLRTGNTDIA--IIVTGNDSSIGA-----RVTSVYFVDKTI-----KGDA 574

QY 208 DQAIRKIMNQAHKVMVHTKELSCKHMEHYKVSSEKRRK 249

Db 575 FNAILPLINNSNAREITNVTSV-----YQVSSSELTALK 608

RESULT 10

B71603
RESA-H3 antigen PFB0915w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000

C:Accession: B71603

R:Cardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;

; Perte, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.

Science 282, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743; PMID:9804551

A:Accession: B71603

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1558 <GAR>

A:Cross-references: GB:AE001424; GB:AE001362; NID:g3845307; PIDN:AACT1972.1; PID:g384530

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0915w

Query Match 8.7%; Score 108.5; DB 2; Length 1558;

Best Local Similarity 26.0%; Pred. No. 30;

Matches 50; Conservative 35; Mismatches 90; Indels 17; Gaps 7;

QY 62 LIEPKIEINKVIPSYQKFEKSLFLQLSSFLERKGVSVQFKDASHPDQIEKALLVL 121

Db 1123 ILEKKIEKHDFKFEAEIKDLEADILKEVSSLEVEEEKLEEV-HELKEVEHII 1181

QY 122 RMDGNV-AILEDIVESDALSEEKVIDMSSGYLNLFVEPKS-EDITHSGIDVSKIAV 179

Db 1182 SGDAHIKLEDDLEEVDDL-KGSILDMKMGDMGDMKESDVTAKUGERVESLKV 1240

QY 180 IERV-----ELRRTNSGGFVPKTFVHRIKETDHDQAIRKIMNQAHKVMVHTKELSKH 234

Db 1241 LSSALGMDDEQMKTKAQRPKLEEVLLKEEVKEPKKTKK---KVRFDIKDKEPKD- 1296

QY 235 MEHYKVSSENMK 246

Db 1297 ---EIVEVENK 1304

RESULT 11

S72378
pd377 protein - Enterococcus faecalis plasmid pPD1

C:Species: Enterococcus faecalis

C>Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 15-Oct-1999

C:Accession: S72378

R:Hirt, H.; Wirth, R.; Muscholl, A.

Mol. Gen. Genet. 252, 640-647, 1996

A:Title: Comparative analysis of 18 sex pheromone plasmids from Enterococcus faecalis: d

A:Reference number: S72375; MUID:97074879; PMID:8917306

A:Accession: S72378

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-377 <HIR>

A:Cross-references: EMBL:X69978; NID:g1272655; PIDN:CAA65679.1; PID:e236362; PID:g127265

A:Experimental source: strain OG1X

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1996

C:Genetics:

A:Gene: pd377

A:Gene: plasmid pPD1

Query Match 8.6%; Score 108; DB 2; Length 377;

Best Local Similarity 22.3%; Pred. No. 5.4;

Matches 56; Conservative 48; Mismatches 79; Indels 68; Gaps 12;

QY 4 GSIAIVIGSLIAGAFYATADGNPAKQHNNTGESVELHFHYPIKQKQEPKNSHLVLI 63

Db 159 GILLIVIGWFKFMAGSQSTAKTSSEPTYQVLVNEKAEIVVKYPEK----- 204

QY 64 EPKIEINKVIPESYQKFEKSLFL-----QLS-----SFLERKGVSVQFKDASHPDQ 112

Db 205 EPDL-----FEELFKEDKAGLIIAHSNTQALYLAFLKDWQKATL---SKLPQE 256

QY 113 IKEKALLVLRMDGNVAILDIVESDALSSEKVIDMSSGYLNLFVPEP-KSEDIHSGFI 171
DB 257 SDVQAW-----GYAFLEGGKEEKLINKE-----IQNVLTQIKSKKEIQAY-- 301
QY 172 DVSKIKAVIER--VELRRTNSGGFVPKTFVHRIKETDHDQAIKIMNQAYHKVMVHTKE 229
DB 302 -----KLLRERKIPAEARIN-----ERLKNELDEAIK--VAKSIHNLLEKYAKD 344
QY 230 LSKKMEHYEK 240
DB 345 KENKELSEKER 355

RESULT 12

T08880
NMDA receptor-binding protein yotiao - human
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T08880
R:Lin, J.W.; Wyszynski, M.; Madhavan, R.; Sealock, R.; Kim, J.U.; Sheng, M.
J. Neurosci. 18, 2017-2027, 1998
A:Title: Yotiao, a novel protein of neuromuscular junction and brain that interacts with
A:Reference number: Z16511; MUID:98151389; PMID:9482789
A:Accession: T08880
A:Status: preliminary;
A:Molecule type: mRNA
A:Residues: 1-1642 <LIN>
A:Cross-references: EMBL:AF026245; NID:g2623067; PIDN:AAB86384.1; PID:g2623068
C:Genetics:
A:Map position: 7q21-22
C:Keywords: brain; cerebral cortex; coiled coil; neuromuscular junction; skeletal muscle

Query Match 8.6%; Score 107.5; DB 2; Length 1642;
Best Local Similarity 20.7%; Pred. No. 38;
Matches 52; Conservative 47; Mismatches 101; Indels 51; Gaps 10;

QY 14 LAGAPYALADGMPAKQCHNNTGESVELHFHYPIKQEPKNSHLVLEPKIEINKVI 73
DB 263 LKNSTHSSAADLQAKQILTHQOOLE-----EQDHL----- 296
QY 74 PPSYQKEFEKSLFLQSSFLER--KGYSVSQPKDA--SEIPQDIKEKALLVLRMDGNVAIL 130
DB 297 -EDYQK--KKEFTQISFLQEKIKYEMEQDKVNSKEIKQEKIIEEL--NTKII 351
QY 131 ED---IVESDALS-----EKVIDMSSGYLNLFVPEKSEDIHSGFIDVSKIKA 178
DB 352 EBEKTLKDKLTADKLLGELQEQVQKQEIKNMKLELTNSKQKQESSEIKQLMG 411
QY 179 VIERVELRRTNSGGFVPKTFVHRIKETDHD-QAIRKIMNQAYHKVMVHTKELSKHVEH 237
DB 412 TVEELQKRNHKSQFETDIVQRMQETQKRLQRLAELENTGQQVQMKQLIKQVNAQ 471
QY 238 YEKVSSEMKKR 248
DB 472 ME-----EMKTR 478

RESULT 13

F89852
extracellular ECM and plasma binding protein [imported] - Staphylococcus aureus (strain
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: F89852
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.: Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F89852
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-340 <KUR>
A:Cross-references: GB:BA000018; PID:g13700680; PIDN:BAB41977.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: ssp

Query Match 8.5%; Score 106; DB 2; Length 340;
Best Local Similarity 25.4%; Pred. No. 6.4;
Matches 63; Conservative 32; Mismatches 69; Indels 84; Gaps 14;

QY 1 MKGSLAIVLGSLLASGAFYALADGMPAKQCHNNTGESVELHFHYPIKQEPKNSHLV 60
DB 1 MKKKLLVLTMTSL-----FATQILNHNHAKA-----SVTESVDKFFVPESG----- 42
QY 61 VLIPEKIEINKVIPESYQKEFEKSLFLQSSFLERKGYSVSQPKDASEIPQDIKEKALLV 120
DB 43 -----INKIIP-AVD-EFKNSPKVNSNLT-DKNFVVSDEK----- 76
QY 121 LRMDGNVAILDIVESDALSSEKVIDMSSGYLNLFVPEKSEDIHSGFIDVSKIKAVI 180
DB 77 -----LNKIVDSAA-----SKIVDK-----NFAVPEK-----LGNIVPEYKEIN 112
QY 181 ERVELRRTN--SGGFVPKTFVHRIKETDHDQAIKIMNQAYHKVMVH--ITKELSKHMEHY 238
DB 113 NRNVNATNPNASQVDKHFVAKGPE-----VNRFTIQ--NKNVHHFITQ-----THY 158
QY 239 EKVSESEMK 246
DB 159 KKVITSYK 166

RESULT 14

A90073
hypothetical protein SA2442 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: A90073
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.: Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: A90073
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-796 <KUR>
A:Cross-references: GB:BA000018; PID:g13702607; PIDN:BAB43747.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2442
C:Superfamily: preprotein translocase secA

Query Match 8.3%; Score 104.5; DB 2; Length 796;
Best Local Similarity 21.9%; Pred. No. 24;
Matches 60; Conservative 43; Mismatches 92; Indels 79; Gaps 13;

QY 16 SGAFYALADGMPAKQCHNNTGESVELHFHYPIKQKQ-----EPKNSHLVLEPKIE 68
DB 538 SSCYISLDDTLVNRWSDSNLAENNQ-----YSLDAQLSQSNLFNRKVKQIVVKAQRIS 594
QY 69 INKVIPESYQKEFEKSLFLQSSFLERKGYSVSQPKDASEIPQDIKEKALLVLRMDGNVA 128
DB 595 EQGVKAREMANEFKESISIQRLDVYEERN-RVLEIDDAEN--RDFKALAKDVPEM----- 646
QY 129 ILEIVESDALSSEKVIDMSSGYLNLFVPEKSEDIHSGFIDVSKI-----KAVI--- 180
DB 647 ----FVNEEKVLTSTRVVEYI--YCNLSF-----QNKDVACVNFKDKQAVVTFL 690
QY 181 -----ERVELRRTNSGG-----FVPKTFVHRI-----KETDHDQAIRKIMNQ----- 217
DB 691 LEQFEKQLALNRKNMQSAYVYVNFQKVLKALDSCWLEQVDYVLQQLKASVNRQNGORN 750

```

Qy 218 ---AYHKVMVHTKELSKCHMEHYKVSSEMKKR 248
Db 751 AIFEYHRVA-----LDSFEVMTNRNKR 773

RESULT 15
B90395
purine NTPase [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: B90395
R:She, O.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aweez, M.J.; Chan-
Jones, I.; Jeffries, A.C.; Kozera, C.O.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, H.
arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: B90395
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-864 <KUR>
A:Cross-references: GB:AF006641; NID:G13815551; PIDN:AAK42417.1; GSPDB:GN00155
C:Genetics:
A:Gene: SS02249

Query Match      8.3%; Score 103.5; DB 2; Length 864;
Best Local Similarity 21.9%; Pred. No. 31;
Matches 64; Conservative 49; Mismatches 84; Indels 95; Gaps 15;

Qy 39 SVELHFH-YPIKGQEPKNSHLVLIIEPKI-EINKVIPESYQ-----KEFEKSL--- 85
Db 283 NIKLFEKYEVLAKSHTEMSANVINLEKEIEEYKAIKREKELEPKYKLEKLEEL 342

Qy 86 -----FLQLSSFLERK-GYSVSQFKDASEIPODI-----KEKALLVLRMDGNVA 128
Db 343 QPKYQYLLKLSDLDSKLNLERLEKDASELSNDIDKVSLEQKVEETRKQLNLAQLA 402

Qy 129 ILEDIVES-----DALSEKVIDMSSGY-----LNLNFVEPK 161
Db 403 KVESLISEKNEIINNISOVEGTCPCVGRPLDEEHKQKIIKEAKSYLQLLELNKNELEE 462

Qy 162 SEDIHSFGIDVSKIKAVIERVELRTNSGGFVPKTFVHRKETDH----- 207
Db 463 LKKITN-----ELAK-----IER-EYRRLSNKASYDNVMQKKNLEENLHSEIESLKN 513

Qy 208 -DQAIRKIMNQA-----YHKVMVHTK-----ELSKKHMEHYKVSSEMKKRK 249
Db 514 IDEEIKKINEEVKELKUYEEFMRLSKYTKBELDKRVRK-----LDEMKKKK 560

```

Search completed: December 10, 2003, 18:37:39
Job time : 21.6217 secs

[illegible]

RESULT 3	HPAA_HELPY	STANDARD;	PRT; 260 AA.
ID	HPAA_HELPY		
AC	P55369;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Neuraminylsucrose-binding hemagglutinin precursor (N-		
DE	acetylneuraminylsucrose-binding fibrillar hemagglutinin receptor-		
DE	binding subunit) (NLBH) (Flagellar sheath adhesin).		
GN	HPAA OR HP0797.		
OS	Helicobacter pylori (Campylobacter pylori).		
OC	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterale		
OC	Helicobacteriaceae; Helicobacter.		
OX	NCBI_TaxID=210;		
[1]			
RP	SEQUENCE FROM N.A.		
RP	STRAIN=26695 / ATCC 700392;		
RC	MEDLINE=97394467; PubMed=9252185;		
RC	Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,		
RA	Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.,		
RA	Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,		
RA	Loftus B., Richardson D., Dodson R., Khakladi H.G., Glodek A.,		
RA	McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,		
RA	Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.N.,		
RA	Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wattley L., Wallin		
RA	Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,		
RA	Wenter J.C.;		
RT	"The complete genome sequence of the gastric pathogen Helicobacter		
RT	pylori.";		
RL	Nature 388:539-547(1997).		
CC	-!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipi		
CC	anchor (Probable).		

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC	EMBL; AE000591; AAD07844.1; -.
CDR	PIR; E64619; E64619.
DR	TIGR; HP0797; -.
DR	Pfam; PF05211; NLBH; 1.
DR	PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR	Flagella; Outer membrane; Lipoprotein; Signal; Complete proteome.
DR	SIGNAL
DR	CHAIN
DR	28
DR	260
DR	NEURAMINYLACTOSE-BINDING HEMAAGGLUTININ
DR	LIPID
DR	28
DR	N-ACYL DIGLYCERIDE (PROBABLE).
DR	SEQUENCE
DR	260 AA; 29040 MW; D9F86C8B0555D071 CRC64;
DR	BY SWIMLARI; 1.
DR	27
DR	28
DR	260
DR	NEURAMINYLACTOSE-BINDING HEMAAGGLUTININ
DR	LIPID
DR	28
DR	N-ACYL DIGLYCERIDE (PROBABLE).
DR	SEQUENCE
DR	260 AA; 29040 MW; D9F86C8B0555D071 CRC64;
DR	BY SWIMLARI; 1.
DR	27
DR	28
DR	260
DR	NEURAMINYLACTOSE-BINDING HEMAAGGLUTININ
DR	LIPID
DR	28
DR	N-ACYL DIGLYCERIDE (PROBABLE).
DR	SEQUENCE
DR	260 AA; 29040 MW; D9F86C8B0555D071 CRC64;
DR	BY SWIMLARI; 1.
DR	27
DR	28
DR	260
DR	NEURAMINYLACTOSE-BINDING HEMAAGGLUTININ
DR	LIPID
DR	28
DR	N-ACYL DIGLYCERIDE (PROBABLE).
DR	SEQUENCE
DR	260 AA; 29040 MW; D9F86C8B0555D071 CRC64;
DR	BY SWIMLARI; 1.
DR	27
DR	28
DR	260
DR	NEURAMINYLACTOSE-BINDING HEMAAGGLUTININ
DR	LIPID
DR	28
DR	N-ACYL DIGLYCERIDE (PROBABLE).
DR	SEQUENCE
DR	260 AA; 29040 MW; D9F86C8B0555D071 CRC64;
DR	BY SWIMLARI; 1.
DR	27
DR	28
DR	260
DR	NEURAMINYLACTOSE-BINDING HEMAAGGLUTININ
DR	LIPID
DR	28
DR	N-ACYL DIGLYCERIDE (PROBABLE).
DR	SEQUENCE
DR	260 AA; 29040 MW; D9F86C8B0555D071 CRC64;
DR	BY SWIMLARI; 1.
DR	27
DR	28
DR	260
DR	NEURAMINYLACTOSE-BINDING HEMAAGGLUTININ
DR	LIPID
DR	28
DR	N-ACYL DIGLYCERIDE (PROBABLE).
DR	SEQUENCE
DR	260 AA; 29040 MW; D9F86C8B0555D071 CRC64;
DR	BY SWIMLARI; 1.
DR	27
DR	28
DR	260
DR	NEURAMINYLACTOSE-BINDING HEMAAGGLUTININ
DR	LIPID
DR	28
DR	N-ACYL DIGLYCERIDE (PROBABLE).
DR	SEQUENCE
DR	260 AA; 29040 MW; D9F86C8B0555D071 CRC64;
DR	BY SWIMLARI; 1.
DR	27
DR	28
DR	260
DR	NEURAMINYLACTOSE-BINDING HEMAAGGLUTININ
DR	LIPID
DR	28
DR	N-ACYL DIGLYCERIDE (PROBABLE).
DR	SEQUENCE
DR	260 AA; 29040 MW; D9F86C8B0555D071 CRC64;
DR	BY SWIMLARI; 1.
DR	27
DR	28
DR	260
DR	NEURAMINYLACTOSE-BINDING HEMAAGGLUTININ
DR	LIPID
DR	28
DR	N-ACYL DIGLYCERIDE (PROBABLE).
DR	SEQUENCE
DR	260 AA; 29040 MW; D9F86C8B0555D071 CRC64;
DR	BY SWIMLARI; 1.
DR	27
DR	28
DR	260
DR	NEURAMINYLACTOSE-BINDING HEMAAGGLUTININ
DR	LIPID
DR	28
DR	N-ACYL DIGLYCERIDE (PROBABLE).
DR	SEQUENCE
DR	260 AA; 29040 MW; D9F86C8B0555D071 CRC64;
DR	BY SWIMLARI; 1.
DR	27
DR	28
DR	260
DR	NEURAMINYLACTOSE-BINDING HEMAAGGLUTININ
DR	LIPID
DR	28
DR	N-ACYL DIGLYCERIDE (PROBABLE).
DR	SEQUENCE
DR	260 AA; 29040 MW; D9F86C8B0555D071 CRC64;
DR	BY SWIMLARI; 1.
DR	27
DR	28
DR	260
DR	NEURAMINYLACTOSE-BINDING HEMAAGGLUTININ
DR	LIPID
DR	28
DR	N-ACYL DIGLYCERIDE (PROBABLE).
DR	SEQUENCE
DR	260 AA; 29040 MW; D9F86C8B0555D071 CRC64;
DR	BY SWIMLARI; 1.
DR	27
DR	28
DR	260
DR	NEURAMINYLACTOSE-BINDING HEMAAGGLUTININ
DR	LIPID
DR	28
DR	N-ACYL DIGLYCERIDE (PROBABLE).

Query Match 15.8%; Score 198.5; DB 1; Length 260;
Best Local Similarity 24.8%; Pred. No. 1.2e-06;
Matches 57; Conservative 65; Mismatches 79; Indels 29; Gaps 10;

39 SVELHFHYPIKQOEPKNSHLVLEPKTEINKVTPESYQKEFEKSFLQLUSSFLERKGY 98
:::| | : : : : : : : : : : : : : : : : : :
39 ALKLNHY-PASEKVQALDEK-ILLRPAFYSDNIATKYENKFNQATALKVEIQLNQGY 96
Ddb

[illegible]

97 KVISVDSSKDDLSFSQ--KKEGYLAVAMNGEIVLRDPKRTTQKKSEFGLLFSFGLDQM 154

144 KVIDMSSGYLNLNFVEPKSEDIHSFGIDVSKIKAVIERVELRRT--NSGGFVPKTFVH 200

155 EGVLPAGFVKVTILEPMGESLDSTMDLSELD--IQEKFLKTHSSHSGGLVST---- 208

10

QY 201 RIKETDH-DOAIRKIMNOYAHKVMVHITKELSKHMEHYKVSSEMCKR 249
 DB 209 MVKGTDSNDAIKSALNKIFANIMQEMDKLTKQNLSEYQKDAKELCKR 258

RESULT 4

ID HPAA HELPJ STANDARD; PRT; 260 AA.
 AC Q9ZL47;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neuraminylactose-binding hemagglutinin precursor (N-acetylneuraminylactose-binding fibrillar hemagglutinin receptor-binding subunit) (NLBH) (Flagellar sheath adhesin).
 DE HPAA OR JHP0733.
 GN Helicobacter pylori J99 (Campylobacter pylori J99).
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxID=85963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C., Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G., Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C., Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F., Trust T.J.;
 RA "Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori."
 RT Nature 397:176-180(1999).
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid anchor (Probable).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X61574; CAA43773.1; -
 CC PIR; C71896; C71896.
 CC PIR; C71896; C71896.
 CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Flagella; Outer membrane; Lipoprotein; Signal.
 FT SIGNAL 1 27 BY SIMILARITY.
 FT CHAIN 28 260 NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ.
 FT LIPID 28 28 N-ACYL DIGLYCERIDE (PROBABLE).
 FT DOMAIN 134 139 N-ACETYL-NEURAMINYL-ALPHA(2,3)-LACTOSE BINDING MOTIF (POTENTIAL).
 SQ SEQUENCE 260 AA; 29199 MW; F5ED538C62B293B8 CRC64;

Query Match 15.5%; Score 194.5; DB 1; Length 260;
 Best Local Similarity 23.9%; Pred. No. 2.3e-06;
 Matches 55; Conservative 67; Mismatches 79; Indels 29; Gaps 10;

QY 39 SVELHFPYPIKQEPKNSHLVLEPKIEINKVIPESYQKEPKSLFLQLSFLERKGY 98
 DB 39 ALKLNH-PASEKVQALDEK-ILLKPAFYSDNIAKEYNKFNQITLKVEILQNGY 96
 QY 99 SV-----SQFKDASEIPQDIKEKALLVLRMDGNVAILED-----IVEESD-----ALSEE 143
 DB 97 KVINVDSSDKDDFSFAQ--KKEGYLAVANNGEIVLRPDPKRTIQKSEPGLLFSTGLDKM 154
 QY 144 KVIDMSSGYLNFVPEKSDIITHSFGIDVSKIKAVIERVELRRT--NSGGFVPKTFVH 200
 DB 155 EGVLIPIAGFKVITILEPMSGESLDSFTMDLSELD--IQEFLKTHSHSGGLVST---- 208

QY 201 RIKETDH-DOAIRKIMNOYAHKVMVHITKELSKHMEHYKVSSEMCKR 249
 DB 209 MVKGTDSNDAIKSALNKIFANIMQEMDKLTKQNLSEYQKDAKELCKR 258

RESULT 5

ID HPAL HELPJ STANDARD; PRT; 260 AA.
 AC Q48254;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Neuraminylactose-binding hemagglutinin precursor (N-acetylneuraminylactose-binding fibrillar hemagglutinin receptor-binding subunit) (NLBH) (Flagellar sheath adhesin).
 GN HPAA.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=8826;
 RX MEDLINE=93139035; PubMed=7678592;
 RA Evans D.G., Karjalainen T.K., Evans D.J., Graham D.Y., Lee C.-H.;
 RT "Cloning, nucleotide sequence, and expression of a gene encoding an adhesin subunit protein of Helicobacter pylori."
 RL J. Bacteriol. 175:674-683(1993).
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid anchor (Probable).
 CC -1- PTM: THE N-TERMINUS IS BLOCKED.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X61574; CAA43773.1; -
 CC PIR; C71896; C71896.
 CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Flagella; Outer membrane; Lipoprotein; Signal.
 FT SIGNAL 1 27 BY SIMILARITY.
 FT CHAIN 28 260 NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ.
 FT LIPID 28 28 N-ACYL DIGLYCERIDE (PROBABLE).
 FT DOMAIN 134 139 N-ACETYL-NEURAMINYL-ALPHA(2,3)-LACTOSE BINDING MOTIF (POTENTIAL).
 SQ SEQUENCE 260 AA; 29166 MW; 22489598065E7B14 CRC64;

Query Match 15.3%; Score 191.5; DB 1; Length 260;
 Best Local Similarity 23.9%; Pred. No. 3.6e-06;
 Matches 55; Conservative 67; Mismatches 79; Indels 29; Gaps 10;

QY 39 SVELHFPYPIKQEPKNSHLVLEPKIEINKVIPESYQKEPKSLFLQLSFLERKGY 98
 DB 39 ALKLNH-PASEKVQALDEK-ILLKPAFYSDNIAKEYNKFNQITLKVEILQNGY 96
 QY 99 SV-----SQFKDASEIPQDIKEKALLVLRMDGNVAILED-----IVEESD-----ALSEE 143
 DB 97 KVINVDSSDKDDFSFAQ--KKEGYLAVANNGEIVLRPDPKRTIQKSEPGLLFSTGLDKM 154
 QY 144 KVIDMSSGYLNFVPEKSDIITHSFGIDVSKIKAVIERVELRRT--NSGGFVPKTFVH 200
 DB 155 EGVLIPIAGFKVITILEPMSGESLDSFTMDLSELD--IQEFLKTHSHSGGLVST---- 208
 QY 201 RIKETDH-DOAIRKIMNOYAHKVMVHITKELSKHMEHYKVSSEMCKR 249
 DB 209 MVKGTDSNDAIKSALNKIFANIMQEMDKLTKQNLSEYQKDAKELCKR 258

RESULT 6

AKA9 HUMAN
 ID AKA9 HUMAN STANDARD; PRT; 3911 AA.
 AC Q99956; O14869; O43355; O94895; O9UOH3; Q9UOQ4; Q9Y6B8; Q9Y6Y2;
 DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE A-kinase anchor protein 9 (Protein kinase A anchoring protein 9)

DE (PKA9) (A-kinase anchor protein 450 kDa) (AKAP 450) (A-kinase anchor

DE protein 350 kDa) (AKAP 350) (hgAKAP 350) (AKAP 120 like protein)

DE (Hyperion protein) (Yotiao protein) (Centrosome- and Golgi-localized

DE PKA-associated protein) (CG-NAP).

GN AKAP9 OR AKAP450 OR AKAP350 OR KIAA0803.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

XP 1

RP SEQUENCE FROM N.A. (ISOFORM 4).

RC TISSUE=Brain;

EX MEDLINE=98151389; PubMed=9482789;

RA Lin J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.;

RT "Yotiao, a novel protein of neuromuscular junction and brain that

RT interacts with specific splice variants of NMDA receptor subunit

RT NR1.";

RL J. Neurosci. 18:2017-2027(1998).

RL 2

RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.

EX MEDLINE=92219864; PubMed=10202149;

RA Witczak O., Skolhegg B.S., Keryer G., Bornens M., Tasken K.,

RA Jahnson T., Oerstavik S.;

RT "Cloning and characterization of a cDNA encoding an A-kinase anchoring

RT protein located in the centrosome, AKAP450.";

RL EMBO J. 18:1858-1868(1999).

RL 3

RP SEQUENCE FROM N.A. (ISOFORM 3).

RC TISSUE=Brain;

EX MEDLINE=99287934; PubMed=10358086;

RA Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.;

RA "Characterization of a novel giant scaffolding protein, CG-NAP, that

RA anchors multiple signaling enzymes to centrosome and the Golgi

RA apparatus.";

RL J. Biol. Chem. 274:17267-17274(1999).

RL 4

RP SEQUENCE FROM N.A. (ISOFORM 1).

EX Kemner W.A., Deiss S., Schwarz U.;

RA "Cloning of Hyperion.";

RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

RL 5

RP SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).

RC TISSUE=Gastric parietal cell;

EX MEDLINE=99115654; PubMed=9915845;

RA Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,

RA Trotter K.W., Milgram S.L., Goldenring J.R.;

RA "AKAP350, a multiply spliced protein kinase A-anchoring protein

RA associated with centrosomes.";

RL J. Biol. Chem. 274:3055-3066(1999).

RL 6

RP SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).

RC TISSUE=Lymphoblast;

EX Hinds K., Sutterer C., Becker M., Hawkins M.;

RA Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

RL 7

RP SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).

RC TISSUE=Lung;

EX Milgram S.L., Goldenring J.R., Schmidt P.H.;

RA "AKAP350: A multiply spliced family of proteins with centrosomal

RA association.";

RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

RL 8

RP SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).

RC TISSUE=Brain;

EX MEDLINE=99087497; PubMed=9872452;

RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,

RA Tanaka A., Kotani H., Nomura N., Ohara O.;

RA "Prediction of the coding sequences of unidentified human genes. XI.

RA The complete sequences of 100 new cDNA clones from brain which code

RA for large proteins in vitro.";

RL DNA Res. 5:277-286(1998).

RL 9

RP SEQUENCE OF 17-1800 FROM N.A.

RA Wu X., Graves T., Bradshaw H.;

RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

CC 1- FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE

CC A. SCAFFOLDING PROTEIN THAT ASSEMBLES SEVERAL PROTEIN KINASES AND

CC PHOSPHATASES ON CENTROSOME AND GOLGI APPARATUS WHERE PHYSIOLOGICAL

CC EVENTS CAN BE REGULATED BY PHOSPHORYLATION STATE OF PROTEIN

CC SUBSTRATES. ISOFORM 4/YOTIAO IS ASSOCIATED WITH THE N-METHYL-D-

CC ASPARTATE RECEPTOR AND IS SPECIFICALLY FOUND IN THE NEUROMUSCULAR

CC JUNCTION (NMJ) AS WELL AS IN NEURONAL SYNAPSES EXPLAINING THAT ITS

CC ROLE MAY BE TO ORGANIZE POSTSYNAPTIC SPECIALIZATIONS.

CC 1- SUBUNIT: INTERACTS WITH THE REGULATORY REGION OF PROTEIN KINASE N

CC (PKN). PROTEIN PHOSPHATASE 2A (PP2A), PROTEIN PHOSPHATASE 1 (PP1)

CC AND THE IMMATURE NON-PHOSPHORYLATED FORM OF PKC EPSILON.

CC 1- SUBCELLULAR LOCATION: CENTROSOMAL IN MANY CELL TYPES AND

CC CYTOPLASMIC IN PARIENTAL CELLS.

CC 1- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=6;

CC Name=1;

CC IsoId=Q99996-1; Sequence=Displayed;

CC Name=2;

CC IsoId=Q99996-2; Sequence=VSP_004102, VSP_004107;

CC Name=3; Synonyms=CG-NAP;

CC IsoId=Q99996-3; Sequence=VSP_004102, VSP_004105, VSP_004107;

CC Name=4; Synonyms=Yotiao;

CC IsoId=Q99996-4; Sequence=VSP_004103, VSP_004104;

CC Name=5;

CC IsoId=Q99996-5; Sequence=VSP_004108;

CC Name=6; Synonyms=AKAP350;

CC IsoId=Q99996-6; Sequence=VSP_004106, VSP_004107, VSP_004109;

CC 1- TISSUE SPECIFICITY: WIDELY EXPRESSED. ISOFORM 4/YOTIAO IS HIGHLY

CC EXPRESSED IN SKELETAL MUSCLE AND IN PANCREAS.

CC 1- DOMAIN: RII BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC HELIX,

CC COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A

CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.

CC 1- CAUTION: REF.6 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO

CC FRAMESHIFTS IN POSITIONS 3782 AND 3811.

CC 1- CAUTION: REF.9 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FOUR

CC FRAMESHIFTS IN POSITIONS 29, 1653, 1699 AND 1735.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch

CC -----

DR EMBL; AJ131693; CAB40713.1; -

DR EMBL; AB019691; BAA78718.1; -

DR EMBL; AJ010770; CAA09361.1; -

DR EMBL; AF026245; AAB86384.1; -

DR EMBL; AF083037; AAD22767.1; -

DR EMBL; AC004013; AAB96867.1; ALT_FRAME.

DR EMBL; AF091711; AAD39719.1; -

DR EMBL; AB018346; BAA34523.1; -

DR EMBL; AC000066; AAC60380.1; ALT_FRAME.

DR Genew; HGNC:379; AKAP9.

DR MIM; 604001; -

DR GO; GO:0005813; C:centrosome; TAS.

DR GO; GO:0005856; C:cytoskeleton; TAS.

DR GO; GO:0004973; F:N-methyl-D-aspartate receptor-associated pr. . . ; TAS.

DR GO; GO:0005515; F:protein binding activity; TAS.

DR GO; GO:0007165; P:signal transduction; TAS.

DR GO; GO:0006832; P:small molecule transport; TAS.

DR GO; GO:0007268; P:synaptic transmission; TAS.

DR Colled coil; Alternative splicing; Polymorphism.

FT DOMAIN 2554 2567 PKA-RII SUBUNIT BINDING DOMAIN.

FT DOMAIN 164 914 COILED COIL (POTENTIAL).

FT DOMAIN 944 1022 COILED COIL (POTENTIAL).

FT DOMAIN 1100 1185 COILED COIL (POTENTIAL).

```
FT DOMAIN 1253 1280 COILED COIL (POTENTIAL).
FT FT 1336 1392 COILED COIL (POTENTIAL).
FT DOMAIN 1434 1459 COILED COIL (POTENTIAL).
FT FT 1585 1659 COILED COIL (POTENTIAL).
FT DOMAIN 1857 2455 COILED COIL (POTENTIAL).
FT FT 2544 2561 COILED COIL (POTENTIAL).
FT DOMAIN 2603 2776 COILED COIL (POTENTIAL).
FT FT 3065 3092 COILED COIL (POTENTIAL).
FT DOMAIN 3124 3470 COILED COIL (POTENTIAL).
FT FT 3587 3689 COILED COIL (POTENTIAL).
FT DOMAIN 3726 3730 POLY-LEU.
FT FT 203 292 GLN-RICH.
FT DOMAIN 321 1010 GLU-RICH.
FT FT 1846 2772 Missing (in isoform 2 and isoform 3).
FT VARSPLIC 17 28 /FTId-VSP_004102.
FT FT VARSPLIC 1637 1642 /FTId-VSP_004103.
FT FT VARSPLIC 1643 3911 Missing (in isoform 4).
FT FT VARSPLIC 2175 2182 /FTId-VSP_004104.
FT FT VARSPLIC 2175 2183 /FTId-VSP_004105.
FT FT VARSPLIC 2895 2907 /FTId-VSP_004106.
FT FT VARSPLIC 2895 2948 (in isoform 2, isoform 3 and isoform 6).
FT FT VARSPLIC 3901 3911 /FTId-VSP_004107.
FT FT VARSPLIC 3901 3911 /FTId-VSP_004108.
FT FT VARSPLIC 3901 3911 /FTId-VSP_004109.
FT VARIANT 1347 1347 K -> Q.
FT FT 76 76 /FTId-VAR_010926.
FT CONFLICT 475 475 E -> Q (IN REF. 3).
FT CONFLICT 554 554 M -> I (IN REF. 3).
FT CONFLICT 638 638 E -> G (IN REF. 3).
FT CONFLICT 663 663 R -> S (IN REF. 3).
FT CONFLICT 913 913 N -> S (IN REF. 3).
FT CONFLICT 913 913 H -> N (IN REF. 3).
FT CONFLICT 956 956 K -> N (IN REF. 3).
FT CONFLICT 980 982 QKH -> PKP (IN REF. 1 AND 2).
FT CONFLICT 997 997 Q -> P (IN REF. 1 AND 2).
FT CONFLICT 1001 1001 Q -> P (IN REF. 1 AND 2).
FT CONFLICT 1020 1020 N -> D (IN REF. 3).
FT CONFLICT 1028 1028 V -> E (IN REF. 3).
FT CONFLICT 1626 1626 R -> P (IN REF. 1 AND 2).
FT CONFLICT 1703 1703 N -> T (IN REF. 3).
FT CONFLICT 1707 1707 V -> G (IN REF. 3).
FT CONFLICT 1802 1803 MISSING (IN REF. 5).
FT CONFLICT 1843 1843 A -> P (IN REF. 3).

Query Match
Best Local Similarity 8.6%; Score 107.5; DB 1; Length 3911;
Matches 52; Conservative 47; Mismatches 101; Indels 51; Gaps 10;

QY 14 LASGAFYALADGMPAKQOHNNNTGESVELHFHYPIKQEPKNSHLVLIIEPKIVKI 73
DB 263 LRNSTHSSAADLQAKQILTHQOOLE-----EQDHL----- 296
QY 74 PESYQKEFEKSLFLQSSFLER--KGYSVQPKDA-SEIPQIKKALVLMDGNVAIL 130
DB 297 -EDYQK--KKEDFTMQISFLQEKIVYEMEQDKVENSNERIQEKTIIEEL--NTRII 351
QY 131 ED---IIVESDALS-----EEKVIDMSSGYLNLNLFVPEKSDIHSFOIDYSKKA 178
DB 352 EEKKTLEKDKLTADKLLGLQELQEQIKVQKQEIKNKLELNSQKQKQSEETKQLNG 411
QY 179 VIERVELRNTSGGVFKTFVHRIKETDHD-QAIRKIMNQAVHYKVMWHITKELSKKMEH 237
DB 412 TVEELQKRNHDSQFETDIVQRMQEQTKQLQRAELDEMYGQIQVQMKQLIQHMAQ 471
QY 238 YEKVSSEMKR 248
```

RESULT 7

```
DB 472 ME---EMKTR 478
ID RASO SULSO STANDARD; PRT; 864 AA.
AC Q97WH0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aweyaz M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erasus G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.,
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
rad50/mre11 complex possesses single-strand endonuclease activity
and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE006829; AAK42417.1; .
CC PIR; B90395; B90395.
CC HAMAP; MF_00449; . 1.
CC InterPro; IPR003439; ABC transporter.
CC InterPro; IPR003405; SMC_C.
CC Pfam; PF04423; Rad50_zn_hook; 1.
CC Pfam; PF02483; SMC_C; 1.
CC Pfam; PF02463; SMC_N; 1.
CC ProDom; PD000006; ABC transporter; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 30 37 ATP (BY SIMILARITY).
FT DOMAIN 164 701 COILED COIL (POTENTIAL).
SQ SEQUENCE 864 AA; 101601 MW; 657076AE9B709FC CRC64;

Query Match
Best Local Similarity 8.3%; Score 103.5; DB 1; Length 864;
Matches 64; Conservative 49; Mismatches 84; Indels 95; Gaps 15;

QY 39 SVELHFT-YPKIQKQEPKNSHLVLIIEPKI-EINKVIPESYQ-----KFEKSL--- 85
DB 283 NIKLKPEKYEVLAKSHTEGANSVINLEKEIEVEKAIKREKELEPKYKYLEKLEEL 342
QY 86 -----FLQSSFLER-KYSVQPKDASEIPDI-----KEKALLVLRMDGNVA 128
DB 343 QPKYQYKLUKSLDLSKLNLERLEKDASELSNDIKVNSLEOKVEETRKQLNLAQLA 402
```

```

QY 129 ILEIVRES-----DALSEKVIDMSSGY-----LNINFEVK 161
DB 403 KVESLISERKEININISQVEGTCVCGRPLDEHKQKIKEAKSYILQELNKELEBE 462
QY 162 SEDIHSGFIDSVKIAVERLRTNSGFGVFKTFVHRIKETDH-----207
DB 463 LKKTIN---ELNK-----IER-EYRLSNKASVDNVRQLKLNBEIENLHSEISLKN 513
QY 208 -DQAIRKMQA-----YHKVMWHITK-----ELSKGMEHYEKSSEMKKRK 249
DB 514 IDEBIKINSEVKELKLYEEMFLSKYTKTELDKRRVK-----LDEMKKKK 560

RESULT 8
TIG AQUAE STANDARD; PRT; 478 AA.
AC O67358;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Trigger factor (TF).
GN TIG OR AQ_1340.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V75;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., AuJay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus";
RL Nature 392:353-358 (1998).
CC -!- FUNCTION: Involved in protein export. Acts as a chaperone by
CC maintaining the newly synthesized protein in an open conformation
CC (By similarity).
CC -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY. TIG SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000735; AAC07311.1; -.
CC FIC; C70416; C70416.
CC HAMAP; MF_00303; -.
CC InterPro; IPR001179; FKBP_PPIase.
CC InterPro; IPR005215; TIG_fac.
CC TIGRfams; TIGR00115; tig; 1.
CC PROSITE; PS00453; FKBP_PPIASE 1; FALSE NEG.
CC PROSITE; PS00454; FKBP_PPIASE 2; FALSE NEG.
CC PROSITE; PS00059; FKBP_PPIASE 3; FALSE NEG.
CC Cell division; Chaperone; Isomerase; Rotamase; Complete proteome.
CC DOMAIN 197 279 PPIASE, FKBP-TYPE.
CC SEQUENCE 478 AA; 54895 MW; 6868FD51F2A5BADO CRC64;

Query Match 8.1%; Score 101.5; DB 1; Length 478;
Best Local Similarity 22.0%; Pred No. 7.9;
Matches 68; Conservative 37; Mismatches 85; Indels 119; Gaps 16;

QY 3 KGSIAVLGSLASGAFYALADGMPAKQOHNTGESVELHFF--HYPIKGQEPKSHLV 60
DB 217 KQETSVLG---QGMLRPEVEEALKGKK-----VGEVEVELKELPLDQEGKEVGK-----263
QY 61 VLIEPKHFI---NKNVIPESQKFEKSLFTQLSSFLERKGYVSQPKDASE-IPQDIKE 115
DB 264 --VNLIKIKIKIKKKVLPE-INDEFKEL-----GY--ASLKDLEEKIREDIKQ 307

```

RESULT 9

VPS5 YEAST

ID VPS5 YEAST STANDARD; PRT; 675 AA.

AC Q92331; O08483;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Vacuolar protein sorting-associated protein VPS5.

GN VPS5 OR GRD2 OR YOR069W OR YOR29-20.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97318765; PubMed=9175702;

RA Notwehr S.F., Hinds A.E.;

RT "The yeast VPS5/GRD2 gene encodes a sorting nexin-1-like protein

RT required for localizing membrane proteins to the late Golgi.";

RL Mol. Biol. Cell 110:1063-1072 (1997).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=97431786; PubMed=9285823;

RA Horadzovsky B.F., Davies B.A., Seaman M.N.J., McLaughlin S.A.,

RA Yoon S., Emr S.D.;

RT "A sorting nexin-1 homologue, Vps5p, forms a complex with Vps17p and

RT is required for recycling the vacuolar protein-sorting receptor.";

RL Mol. Biol. Cell 8:1529-1541 (1997).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=97279235; PubMed=9133743;

RA Valens M., Bohn C., Daignan-Fornier B., Dang V., Bolotin-Fukuhara M.;

RT "The sequence of a 54.7 kb fragment of yeast chromosome XV reveals

RT the presence of two tRNAs and 24 new open reading frames.";

RL Yeast 13:379-390 (1997).

CC -!

FUNCTION: REQUIRED FOR RETENTION OF LATE GOLGI MEMBRANE PROTEINS

AND VACUOLAR BIOGENESIS.

CC -!- SUBUNIT: INTERACTS WITH VPS17.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- PTM: PHOSPHORYLATED ON SERINE RESIDUE(S).

CC -!- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.

CC -!- SIMILARITY: Contains 1 Phox homology (PX) domain.

CC -!- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A

FRAMESHIFT IN POSITION 13.

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See http://www.isb-sib.ch/announce/

or send an email to license@isb-sib.ch).

CC

EMBL; U73512; AAB62976.1; -.

EMBL; U94735; AAB41798.1; -.

```

DR EMBL; Z70678; CAA94554.1; ALT_FRAME.
DR EMBL; Z74977; CAA99262.1; -.
DR SGD; S0005595; VPSS.
DR InterPro; IPR001683; PX.
DR Pfam; PF00787; PX; 1.
DR SMART; SM00312; PX; 1.
DR PROSITE; PS00195; PX; 1.
DR Transports; Protein transport; Golgi stack; Phosphorylation.
FT DOMAIN 279 394
SQ SEQUENCE 675 AA; 76484 MW; FBA9994EADBCD2BD CRC64;

Query Match
Best Local Similarity 26.6%; Pred. No. 12;
Matches 47; Conservative 35; Mismatches 56; Indels 39; Gaps 11;

QY 96 KGY-----SVSQFKDASETPQDIKEKALLVLRMDGNVAILEDIVESD-----ALSEE 143
DB 445 KGFMSISFSLPKYNEADFFIEKKQK---IDLEDNLKIKSLSEWVDTSRNTLAASTE 501
QY 144 KVTDMSSGYLNLNFVSPKSEDIHISFGIDYSK-IKAVIERVELRRITNSGGFVPTFVHRI 202
DB 502 EFSMVTETLASLVNSPENSE-LNNFA-DVHKSIKSLERSSLGQETLTGVMLDDYIRSL 559
QY 203 KETDHDQAIRKINQV-----AYHKVMVHITKLSKHE-----HYEKVSSEMKK 247
DB 560 -----ASVKAIFNQRSLGY--FLVVIENDMNKXHSQKLGKLNHSEKF-REMEK 607

RESULT 10
SLP1_CABEL STANDARD; PRT; 576 AA.
AC P34260;
DT 01-FEB-1994 (Rel. 28, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein slp-1.
GN SLP-1 OR B0303.9.
OC Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=92168156; PubMed=1538779;
RA Halloran N., Green P., Thierry-Mieg J., Qiu L., Dear S., Coulson A.,
RA Craxton M., Durbin R., Berks M., Metzstein M., Hawkins T.,
RA Ainscough R., Waterston R.;
RT "The C. elegans genome sequencing project: a beginning.";
RL Nature 356:37-41(1992).
CC -!- SIMILARITY: TO YEAST SLP1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; M77697; AAA27901.1; -.
DR PIR; S27790; S27790.
DR WormPep; B0303.9; CE000012.
DR InterPro; IPR001619; Sec1-like.
DR Pfam; PF00995; Sec1; 1.
DR Protein transport.
KW Protein transport.
SQ SEQUENCE 576 AA; 64927 MW; 811FC4A06D896C9A CRC64;

Query Match
Best Local Similarity 8.0%; Score 100.5; DB 1; Length 576;
Matches 47; Conservative 39; Mismatches 63; Indels 71; Gaps 11;

QY 55 KNSHLVLIIEPKIEINKVIPESYQKEFKSLFLQLSSFLERKGYVSQFKD-ASEIPODI 113
DB 294 KHSHINAV---SIEASKVLAIIRDDE-----QFDRDKMSVAEYSLVKCPKII 339
QY 114 KEKALLVLRM-----DGNVAILEDIVESDASEEKVIDMSSGYLNLNFE 159
DB 340 NRKGMIEVHMRLAEMIQSHVYCKOSDSIKJERDLLEYSDS---DKAI 383
QY 160 PKSEDIHISFGIDYSKIKAVIERVELRRITNSGGFVPTFVHRIKETDHDQAIRKINQV 219
DB 384 PLIEDLI-----FDASPLNAVLRLLISVHSLTCTGGLKPSVLQH-----YRRVINGSY 429
QY 220 -----HKVM-----VHITKEL---SKKMEHYEKVSSEMKK 247
DB 430 OSSALNKVLYKQRMGLIREKGGGKMQCEYAAQMMFQOMKK 469

RESULT 11
RASO_PYRAB STANDARD; PRT; 880 AA.
AC Q9UUC8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR PYRAB12200 OR PAB0812.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GES / Orsay;
RX PubMed=12822808;
RA Cohen G.N., Barbe V., Flament D., Galperin M., Hellig R., Lecompte O.,
RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
RT archaeon Pyrococcus abyssi.";
RL Mol. Microbiol. 47:1495-1512(2003).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AJ248286; CAB50131.1; -.
DR PIR; F75103; F75103.
DR HANAP; MF_00449; 1.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR003395; SMC N.
DR Pfam; PF04423; Rad50_zn_hook; 1.
DR Pfam; PF02463; SMC N; 1.
DR ProDom; PD000006; ABC transporter; 1.
DR SMART; SM00382; AAA; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 30 37
FT DOMAIN 144 745
SQ SEQUENCE 880 AA; 103970 MW; FDB177EC7E026479 CRC64;

Query Match
8.0%; Score 100; DB 1; Length 880;

```

Best Local Similarity 20.0%; Pred. No. 20;
Matches 54; Conservative 57; Mismatches 73; Indels 86; Gaps 14;

QY 47 PIKQKQPKSHLVVL-----IEPKINKVIPSQKPEKSLFLOLSSPLERKGY 99
DB 222 PIRGELEKVENKVESIKGKISLQVVKL--KGRKGLBEKI-VQIERSIEKKAK 278
QY 100 VSQPKDASEIPQDI-----KEKALLVL-----RMDGNVAILEDIVE 135
DB 279 ISELE--EIVQIPKQKEKEKVKLGRDEYVESKRLERLEKSKWSELKAIEVVK 335
QY 136 ESDALSE-----EKVDMSSGYLNLFVPEKSEDIHSPGIDVSKIKAVIERVELRRTN 189
DB 336 EGEKKKRAEIREKLEIK--RLELRPYVEEL-----EBAQVQKQIEKRLK- 386
QY 190 SGGFVPKTFVHRI-----KETDHDQAIRKI-----MNQAYHKVMV 224
DB 387 --GLSPGEVIEKLEKERTETEIEAIKEITTRIGQMEQKERNKRAIELRKAIGKCPV 444
QY 225 ---HITKELSKHNEY-----EKUSSEMKK 247
DB 445 CGRELTEHKKELMERYTLEIKKIEBELK 474

RESULT 12
ACVS_EMENI STANDARD; PRT; 3770 AA.
ID ACVS_EMENI
AC P27742;
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE N-(5-amino-5-carboxypentanoyl)-L-cysteiny-D-valine synthase
DE (EC 6.3.2.26) (Delta-(L)-alpha-aminoacyl)-L-cysteiny-D-valine
DE synthetase) (ACV synthetase) (ACVS).
GN ACVA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=G191;
RX MEDLINE=91286299; PubMed=2061333;
RA Maccabe A.P., van Liemt H., Pallissa H., Unkles S.E., Riach M.B.R.,
RT Pfeifer E., von Doehren H., Kinghorn J.R.;
RT "Delta-(L)-alpha-aminoacyl)-L-cysteiny-D-valine synthetase from
RT Aspergillus nidulans. Molecular characterization of the acva gene
RT encoding the first enzyme of the penicillin biosynthetic pathway.";
RL J. Biol. Chem. 266:12646-12654(1991).
CC -!- FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPPTIDE
CC ACV ARE ACTIVATED AS AMINOACYL-ADENYLATES WITH PEPTIDE BONDS
CC FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER
CC INTERMEDIATES.
CC -!- CATALYTIC ACTIVITY: L-2-aminohexanedioate + L-cysteine + L-valine
CC + 3 ATP = N-[L-5-amino-5-carboxypentanoyl]-L-cysteiny-D-valine +
CC 3 AMP + 3 diphosphate.
CC -!- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTHETHEINES
CC (POTENTIAL).
CC -!- PATHWAY: Biosynthesis of penicillin and cephalosporin; first step.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -!- SIMILARITY: Contains 3 acyl carrier domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X54853; CAA38631.1; -.

PIR; A40889; A40889.
HSP; P14687; LAMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR006163; Pp_bind.
DR InterPro; IPR006162; Pplantn attach.
DR InterPro; IPR000379; Ser estrs site.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00501; AMP-binding; 3.
DR Pfam; PF00668; Condensation; 3.
DR Pfam; PF00550; pp-binding; 3.
DR Pfam; PF00975; Thioesterase; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 3.
DR PROSITE; PS00455; AMP BINDING; 3.
DR PROSITE; PS00075; ACP_DOMAIN; 3.
KW Ligase; Antibiotic biosynthesis; Multifunctional enzyme;
KW Repeat; Phosphopantetheine.
FT REPEAT 321 910 DOMAIN 1 (ADIPATE-ACTIVATING).
FT REPEAT 1413 1993 DOMAIN 2 (CYSTEINE-ACTIVATING).
FT REPEAT 2494 3078 DOMAIN 3 (VALINE-ACTIVATING).
FT DOMAIN 850 919 ACYL CARRIER (ACP) 1.
FT DOMAIN 1929 2002 ACYL CARRIER (ACP) 2.
FT DOMAIN 3020 3087 ACYL CARRIER (ACP) 3.
FT BINDING 882 882 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 1965 1965 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 3050 3050 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT ACT SITE 3623 3623 THIOESTERASE (BY SIMILARITY).
SQ SEQUENCE 3770 AA; 422448 MW; CB66B6D232A58CB0 CRC64;
Query Match 7.9%; Score 98.5; DB 1; Length 3770;
Best Local Similarity 20.7%; Pred. No. 1.4e+02;
Matches 57; Conservative 42; Mismatches 114; Indels 63; Gaps 12;
QY 20 YTLADGMP--AKQCHNNTGESV-ELHPHYPIKQK-----EP-----K 55
DB 479 YISGTGFPKGLKQHTNVNSITDLSARYGTGDHHEAILLSAYVPEFVRQMLMALV 538
QY 56 NSHLVLVIE--PKIINKVIPSQKPEKSLFLOLSSPLERKGY-SVSQPKDASEIPQD 112
DB 539 NGHLLAMVDRAEKYDAEKLIIP-FIREHKITLNGTASVLQEDYDFSSCPSLKLILVGEN 596
QY 113 IREKALLVIRMDGNVAILEDIVESDALSSEKVIDMSSGYLNLFVPEKSEDIHSPGID 172
DB 597 LTESYLAIRHRFKNCILNEY-----GFTSAFVTA-----LNVPEGSARNTSLGRP 645
QY 173 VSKIKAVIERVELRRT-----NSGGFVPKTFVHRIKETDHDQAIR 212
DB 646 VENVKCYILNKLKRVPIGATGELHIGLIGSKYLNLRPDLPQRFIPNPFQDHEKELG 705
QY 213 KIMNQAYHKVMVHTKELSKQKHEHYKVSSEMKK 248
DB 706 --LNQLMYKT-GDLARWLPNGEIEYLGRADFQIKLR 738
RESULT 13
FRIL_SOYBN STANDARD; PRT; 250 AA.
ID FRIL_SOYBN
AC P19876;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ferritin 1, chloroplast precursor (SOF-35) (SferH-1).
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91376116; PubMed=1896472;
RA Lescure A.-M., Proudhon D., Pessey H., Ragland M., Theil E.C.,
RA Briat J.-F.;


```

RT "Ferritin gene transcription is regulated by iron in soybean cell
RL cultures.";
RN Proc. Natl. Acad. Sci. U.S.A. 88:8222-8226(1991).
RP [2]
RX SEQUENCE OF 1-189 FROM N.A.
RC TISSUE=HYPOCOTYL;
RD MEDLINE=91009336; PubMed=2211706;
RE Ragland M., Briat J.-F., Gagnon J., Laulhere J.-P., Massenet O.,
RA Theil E.C.;
RR "Evidence for conservation of ferritin sequences among plants and
RT animals and for a transit peptide in soybean.";
RL J. Biol. Chem. 265:18339-18344(1990).
RN [3]
RP SEQUENCE OF 50-63.
RC STRAIN=cv. Mandarin; TISSUE=Seed;
RD MEDLINE=91030706; PubMed=2264818;
RE Lessure A.-M., Massenet O., Briat J.-F.;
RA "Purification and characterization of an iron-induced ferritin from
RT soybean (Glycine max) cell suspensions.";
RL Biochem. J. 272:147-150(1990).
RN CC -!- FUNCTION: Ferritin is an intracellular molecule that stores iron
CC in a soluble, nontoxic, readily available form. The functional
CC molecule, which is composed of 24 chains, is roughly spherical and
CC contains a central cavity into which the polymeric ferric iron
CC core is deposited.
CC CC -!- SUBCELLULAR LOCATION: CHLOROPLAST AND OTHER PLASTIDS.
CC CC -!- TISSUE SPECIFICITY: Leaf > HYPOCOTYL.
CC CC -!- MISCELLANEOUS: MULTIPLE CLEAVAGE SITES MAY OCCUR IN THE EP
CC YIELDING SEVERAL SMALLER (26.5 kDa) FERRITIN SUBUNITS.
CC CC -!- SIMILARITY: BELONGS TO THE FERRITIN FAMILY.
CC CC -!- SIMILARITY: Contains 1 ferritin-like diiron domain.
CC CC -!- CAUTION: Ref.1 n-terminal sequence differs from that shown due to
CC vector contamination.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensese@isb-sib.ch).
CC -----
DR EMBL; M58336; AAA33958.1; ALT_SEQ.
DR EMBL; M72894; AAA34016.1; ALT_SEQ.
DR EMBL; M64337; AAA33953.1; -.
DR PIR; A40992; A40992.
DR HSSP; P07229; 1BG7.
DR InterPro; IPR001519; Ferritin.
DR Pfam; PF00210; ferritin; 1.
DR Prodom; PD000971; Ferritin; 1.
DR PROSITE; PS00540; FERRITIN_1; 1.
DR PROSITE; PS00204; FERRITIN_2; 1.
DR PROSITE; PS0905; FERRITIN_LIKE; 1.
DR Iron storage; Iron; Metal-binding; Chloroplast; Transit peptide.
DK TRANSIT 1 49 CHLOROPLAST.
FT CHAIN 1 250 FERRITIN 1..
FT DOMAIN 50 82 EXTENSION PEPTIDE (EP).
FT FT DOMAIN 83 236 FERRITIN-LIKE DIIRON.
FT METAL 100 100 IRON (BY SIMILARITY).
FT METAL 134 134 IRON (BY SIMILARITY).
FT METAL 135 135 IRON (BY SIMILARITY).
FT METAL 137 137 IRON (BY SIMILARITY).
FT METAL 138 138 IRON (BY SIMILARITY).
FT METAL 184 184 IRON (BY SIMILARITY).
FT CONFLICT 161 161 A -> V (IN REF. 2).
FS SEQUENCE 250 AA; 28050 MW; 3A48F00B33D7CBA9 CRC64;

Query Match 7.8%; Score 98; DB 1; Length 250;
Best Local Similarity 26.8%; Pred.No. 6.4;
Matches 48; Conservative 27; Mismatches 62; Indels 42; Gaps 10;
QY 52 QEPKNSHLVLLIEPKI-----EINKVIPESYOKEFEKS-LFLQLGSFLER----K 96

```

QY 209 -----QAIRKIMNQAY-----HKVMVHTKELSKKH 234
 Db 204 PGGSEITVQKINWVAQNGYISVKSIDSRLRWANIVKARHVAMFLTGTILKKN 255

RESULT 15

GYRB_HAEIN STANDARD; PRT; 806 AA.
 AC P43701;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA gyrase subunit B (EC 5.99.1.3).
 GN GYRB OR H10567.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=rd / KW20 / ATCC 51907;
 RX MEDLINE=9530630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kervanage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uytterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 RL Science 269:496-512 (1995).
 CC -!- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
 CC STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE
 CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
 CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
 CC -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
 CC of double-stranded DNA.
 CC -!- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
 CC BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE
 CC ENZYME FORMS AN A2B2 TETRAMER.
 CC -!- SIMILARITY: Belongs to the type II topoisomerase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U32738; AAC22225.1; -;
 DR F1R; A64078; A64078.
 DR HSSP; P06982; 1AU6.
 DR TIGR; H10567; -;
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR002288; DNA_gyraseB_C.
 DR InterPro; IPR000565; DNA_gyrb.
 DR InterPro; IPR001241; DNA_topoisom.
 DR InterPro; IPR006171; Toprim_dom.
 DR Pfam; PF00204; DNA_gyraseB; 1.
 DR Pfam; PF00986; DNA_gyraseB_C; 1.
 DR Pfam; PF02518; HATPase_C; 1.
 DR Pfam; PF01751; Toprim; 1.
 DR PRINTS; PS00418; TPI2FAMILY.
 DR ProDom; PD149633; DNA_gyraseB_C; 1.
 DR SMART; SM00387; HATPase_C; 1.
 DR SMART; SM00433; TOP2c; 1.
 DR TIGRFAMs; TIGR01059; gyrb; 1.
 DR PROSITE; PS00177; TOPOISOMERASE_II; 1.

KW Topoisomerase; Isomerase; ATP-binding; Complete proteome.
 SQ SEQUENCE 806 AA; 89884 MW; 0E9EC804D134B092 CRC64;

Query Match 7.8%; Score 98; DB 1; Length 806;
 Best Local Similarity 21.8%; Pred. No. 25;
 Matches 52; Conservative 46; Mismatches 68; Indels 72; Gaps 15;

QY 17 GAFYALADGMPAKQOHNNTGESVELHFFHYPIKQBPKNHVLVLEPKIEINKV----72
 Db 471 GLTITLALCGGI-GRDEVNPD-----KLRYH-----HIIIMDADVGGSHITLL 513
 QY 73 -----IPESVQKEFEKSLFLQLSSFLERKGYSVQFQDASEIPQDIKEKALVLRMDG 125
 Db 514 LTFYVRQMPLELIERGY--VYIAQPLLYKVKKGQERYIKQADEN-----EYELTTLALDG 566
 QY 126 -----NVAILEDIVEESDALSEEKVIDMSSGYLNLNFEVPEKSEDIHS--FGI 171
 Db 567 AELHISTNAPAMNALVPEKLVAEYN--SVQKLI-----GRLNRHYPAVPVLOGLIYQSPISI 620
 QY 172 DVSKIKAVIERVELRRITNSGGFVPTFVHRI--KETD-HDQAIRKIMN---QAYHKVM 223
 Db 621 EMKKEESAVE-----NWG-----KSFVEQLTAKETAHQYSVRTQFNAERQVVEAVI 667

Search completed: December 10, 2003, 18:34:30
 Job time : 15.7775 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:29:51 ; Search time 36.1767 Seconds
(without alignments)
1776.146 Million cell updates/sec

Title: US-10-080-113-1
Perfect score: 1254
Sequence: 1 MKGSLAIVLGSLLASGAFY.....LSKKMEHYEKVSSEMKRK 249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1254	100.0	249	16 O25166	O25166 helicobacte
2	1199	95.6	249	16 Q9ZKG5	Q9ZKG5 helicobacte
3	302.5	24.1	278	16 O25234	O25234 helicobacte
4	300	23.9	282	16 Q9ZLY2	Q9ZLY2 helicobacte
5	201.5	16.1	260	2 Q8RNU4	Q8RNU4 helicobacte
6	132	10.5	207	16 O25713	O25713 helicobacte
7	129	10.3	207	16 Q9ZM78	Q9ZM78 helicobacte
8	127.5	10.2	5458	5 Q9U459	Q9U459 plasmodium
9	127.5	10.2	5507	5 Q8IHN3	Q8IHN3 plasmodium
10	116.5	9.3	16215	5 Q9NFS3	Q9NFS3 drosophila
11	116.5	9.3	18074	5 Q917U4	Q917U4 drosophila
12	115.5	9.2	722	16 Q926X5	Q926X5 listeria in
13	113.5	9.1	535	4 Q9NWC6	Q9NWC6 homo sapien
14	113.5	9.1	1416	4 Q9HCL1	Q9HCL1 homo sapien
15	113.5	9.1	1416	4 Q9BZF9	Q9BZF9 homo sapien
16	113.5	9.1	1449	6 Q9BG87	Q9BG87 bos taurus

ALIGNMENTS

RESULT 1

O25166 PRELIMINARY; PRT; 249 AA.
 ID O25166
 AC O25166;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Putative neuraminyllactose-binding hemagglutinin homolog (HPAA).
 GN HP0410.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26595 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185;
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Utterback R.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA "The complete genome sequence of the gastric pathogen Helicobacter
 pylori".
 RT Nature 388:539-547(1997).
 RL Nature 388:539-547(1997).
 DR EMBL; AB000557; AAD07478.1; -.
 DR TIGR; HP0410; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 249 AA; 28349 MW; 43E6ABBA6EABBS15 CRC64;

Query Match 100.0%; Score 1254; DB 16; Length 249;
 Best Local Similarity 100.0%; Pred. No 7.2e-80;
 Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKGSLAIVLGSLLASGAFYALADGMPAKQCHNNTGESVELHPHYFKQEPKNSHLV 60

```

Db 1 MKKSLAIVLGSLLASGAFYFALADGMPAKQHNNTGESVELHFHYPIKQKQEPKSHLV 60
QY 61 VLIEPKIEINKVIPESYQKEFEKSLFLQSLSFLEKRGYSVSQFKDASEIPQDIKEKALLV 120
Db 61 VLIEPKIEINKVIPESYQKEFEKSLFLQSLSFLEKRGYSVSQFKDASEIPQDIKEKALLV 120
QY 121 LRMDGNVAILEDIVESDALSEEKVIDMSSGYLNFVPEKSEDIHSPGIDVSKIKAVI 180
Db 121 LRMDGNVAILEDIVESDALSEEKVIDMSSGYLNFVPEKSEDIHSPGIDVSKIKAVI 180
QY 181 ERVELRRTNSGGFVPKTFVHRIKETDHDQAIRKIMNQAYHKVMVHITKELSKKMEHYEK 240
Db 181 ERVELRRTNSGGFVPKTFVHRIKETDHDQAIRKIMNQAYHKVMVHITKELSKKMEHYEK 240
QY 241 VSSEMKKRRK 249
Db 241 VSSEMKKRRK 249

RESULT 2
Q9ZKGS PRELIMINARY; PRT; 249 AA.
AC Q9ZKGS;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE Putative PARALOG of HPAA.
GN JHP0971.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.B., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180 (1999).
DR EMBL; AE001526; AAD06549.1; -
KW Complete proteome.
SQ SEQUENCE 249 AA; 28513 MW; EC8D9AFDC99548AA CRC64;

Query Match 35.68; Score 1199; DB 16; Length 249;
Best Local Similarity 94.8; Pred. No. 4.9e-76;
Matches 236; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 1 MKKSLAIVLGSLLASGAFYFALADGMPAKQHNNTGESVELHFHYPIKQKQEPKSHLV 60
Db 1 MKKSLAIVLGSLLASGFTTALADGMPKQHNNGESVELHFHYPIKQKQEPKSHLV 60
QY 61 VLIEPKIEINKVIPESYQKEFEKSLFLQSLSFLEKRGYSVSQFKDASEIPQDIKEKALLV 120
Db 61 VLIDPKIEANKVIPENYQKEFEKSLFLQSLSFLEKRGYSVSQFKDASEIPQDIKEKALLV 120
QY 121 LRMDGNVAILEDIVESDALSEEKVIDMSSGYLNFVPEKSEDIHSPGIDVSKIKAVI 180
Db 121 LRMDGNVAILEDIVESDALSEEKVIDMSSGYLNFVPEKSEDIHSPGIDVSKIKAVI 180
QY 181 ERVELRRTNSGGFVPKTFVHRIKETDHDQAIRKIMNQAYHKVMVHITKELSKKMEHYEK 240
Db 181 ERVELRRTNSGGFVPKTFVHRIKETDHDQAIRKIMNQAYHKVMVHITKELSKKMEHYEK 240
QY 241 VSSEMKKRRK 249
Db 241 VSSEMKKRRK 249

```

```

RESULT 3
Q25234 PRELIMINARY; PRT; 278 AA.
AC Q25234;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein HP0492.
GN HP0492.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-P., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Loftus B., Richardson D., Dodson R., Kirkness E.F., Peterson S.,
RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547 (1997).
DR EMBL; AE000564; AAD07568.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 278 AA; 31947 MW; 32F8F72010FC20E4 CRC64;

Query Match 24.1; Score 302.5; DB 16; Length 278;
Best Local Similarity 29.1; Pred. No. 1.8e-13;
Matches 73; Conservative 57; Mismatches 110; Indels 11; Gaps 5;

QY 4 GSLAIVLGSLLASGAFYFALADGMPAKQHNNTGESVE---LHFHYPIKQKQEPKSHL 59
Db 21 GLSSVLICAMNPSAETKTPNDKKNQVQTHRMKTSSEHTPLDFNPIHIVQAPQNHV 80
QY 60 VLIEPKIEINKVIPESYQKEFEKSLFLQSLSFLEKRGYSVSQFKDASEIPQDIKEKALL 119
Db 81 VGILTPRIQVSDNL-KPYIDKFQDALINQITPEKRGYQVLRFDQEKALNAQDKRKIFS 139
QY 120 VLRMDGNVAILEDI---VESDALSEEKVIDMSSGYLNFVPEKSEDIHSPGIDVSKI 176
Db 140 VLDLKGWVGIILEDLKNLKNLPNNPNDLTLDVQSSGVSNFNFYEPESNRVHDFAVEVGT 199
QY 177 KAVIERVELRRTNSGGF-VPKTFVHRIKETDHDQAIRKIMNQAYHKVMVHITKELSKKH 235
Db 200 QAM--TYTKHNSSGLNNSNSIIEHYLEKNKEDAIHKILNRMYAVVMKAVTELTKENI 257
QY 236 EHYEKVSEMK 246
Db 258 DKYREADRMRK 268

RESULT 4
Q9ZLY2 PRELIMINARY; PRT; 282 AA.
AC Q9ZLY2;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE Putative PARALOG of HPAA.
GN JHP0444.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]

```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen W., Mills D.M., Ives C.,
RA Gibson R., Werberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
DR EMBL; AE001479; AAD06033.1; -.
KW Complete proteome.
SQ SEQUENCE 282 AA; 32075 MW; 555653A38461D7A9 CRC64;

Query Match      23.9%; Score 300; DB 16; Length 282;
Best Local Similarity 29.7%; Pred. No. 2.7e-13;
Matches 76; Conservative 57; Mismatches 107; Indels 16; Gaps 6;

QY      6  LAIVGLSLASGAFYALADGMP--AKQQ-----HNTGESVE----LHFHYPIKGKQEP 54
Db      18  VAUGLSSVLIGCANPFAETKFNDAKNOQPVQVOTHERMTTSEHVTPDFNYPVHIVQAP 77

QY      55  KNSHLVLIPEKIEINKVPIESYQKEFEKSLFLQLSFLERKGYVSQFKDASEIPQDIK 114
Db      78  QNHVVVGIEMPRIQVSDNL-KPYIDKFDALINQIOTIFEKGYQVLRFODEKALNVQDK 136

QY      115  EKALLVLRMDCNVAILEDI---VEESDALSSEKVIDMSSGYLNFVEPKSEDIHSPGI 171
Db      137  KKIFSVDLKGWVGILEDLKNLKNLDPNSPLDTVDQSSGVWFNFYEPESRVVHDFAV 196

QY      172  DVSKIKAVIERVELRTNSGGF-VPKTFVRIKETDHDQAIRKMQAYHKVWHITKEL 230
Db      197  EVCTFQAITVYVSTNNSAGGFNSKSVIHENLDKNEBDAIHKLNRVAVVWKKAVTEL 256

QY      231  SKKHMEHYKVSSEM 246
Db      257  TKENIARYDAIDRMK 272

RESULT 5
Q8RNU4 PRELIMINARY; PRT; 260 AA.
AC Q8RNU4;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Adhesin protein HpaA.
GN HPAa.
OS Helicobacter pylori.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH-CTX1;
RX MEDLINE=20123532; PubMed=10660136;
RA Opazo P., Muller I., Rollan A., Valenzuela P., Yudelevich A.,
RA Garcia-de la Guardia R., Urra S., Venegas A.;
RT "Serological response to Helicobacter pylori recombinant antigens in
RT Chilean infected patients with duodenal ulcer, non-ulcer dyspepsia and
RT gastric cancer."
RL APWIS 107:1069-1078(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CH-CTX1;
RA Muller I., Medina-Selby A., Palacios J.L., Martinez P., Opazo P.,
RA Bruce E., Mancilla M., Valenzuela P., Yudelevich A., Venegas A.;
RT "Cloning and comparison of ten gene sequences of a Chilean
RT Helicobacter pylori strain with other Helicobacter pylori strains
RT revealed higher amino acid sequence variability for VacA and CagA
RT virulence factors."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF479028; AAL68698.1; -.
```

```
SQ SEQUENCE 260 AA; 29129 MW; A539FBA85E3376A1 CRC64;

Query Match      16.1%; Score 201.5; DB 2; Length 260;
Best Local Similarity 24.8%; Pred. No. 1.8e-06;
Matches 57; Conservative 65; Mismatches 79; Indels 29; Gaps 10;

QY      39  SVELHFYPIKGKQEPKNSHLVLIPEKIEINKVPIESYQKEFEKSLFLQLSSFLERKGY 98
Db      39  AKLNYH-PASEKVALDER-ILLRPAFYSDNIAKEYENKFNQATKALKEQILQNGQY 96

QY      99  SV---SQFKDASEIPQDIKEKALLVLRMDCNVAILEDI---IVEESD-----ALSEE 143
Db      97  KVINVDSSDKDDPFAQ--KKEGYLAVANGEIVLRPDKETIQKSEPGLLFTGLDKM 154

QY      144  KVIDMSSGYLNFVEPKSEDIHSPGIDVSKIKAVIERVELRT---NSGGFVPKTFVH 200
Db      155  EGVLIAPAGIKVITILEPMWGESLDSFTWDLSELD--IQEKFLKTHSHSGGLVST---- 208

QY      201  RIKETDH-DQAIRKMQAYHKVWHITKELSKKHMEHYKVSSEMKKK 249
Db      209  MVKGTDSNDAIKGALNKAFINIMQEIQDKLTKQNLSYQDKAKELNKR 258

RESULT 6
Q25713 PRELIMINARY; PRT; 207 AA.
AC Q25713;
DT 01-JAN-1998 (T-EMBLrel. 05, Created)
DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Hypothetical protein HP1081.
GN HP1081.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tombs J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G., B.A.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty S.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
DR EMBL; AE000615; AAD08133.1; -.
DR TIGR; HP1081; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 207 AA; 23737 MW; 9BF3A0E76590AB37 CRC64;

Query Match      10.5%; Score 132; DB 16; Length 207;
Best Local Similarity 25.9%; Pred. No. 0.096;
Matches 55; Conservative 35; Mismatches 64; Indels 58; Gaps 11;

QY      42  LHFHYPIKGKQEPKNSHL---VLIPEKIEIN-KVPIESYQKEFEKSLFLQLSSFLERK 97
Db      24  LHLKY-----KDYPKNSALKTAFTLTTPPKIFFNARFVPFPYQKEFKKAITQOIAYFL 75

QY      98  YSVSQFKDASEIPQDIKEKALLVLRMDCNVAIL-LEDIVEESDALSE--EKVIDMSSGYL 154
Db      76  -----KKSAFILNVSGNVFFSEENPKDLKAKELKKTIEPNA----- 115

QY      155  LNFVEPKSEDIHSPGIDVSKIKAVIER-----VELRTNSGGFVPKTFV 199
Db      116  ---DPKA--VVRFLNLQASLILECVPTTCFFDTLLIPTAFSPVPTVYANRLGNPSLFS 169
```

QY 200 HRIKETHDQAIKIMNOAYHKVAVHITKELS 231
 DB 170 QEDK-TYHN-ALIKALNKAYISLMEGLEKRLN 199

RESULT 7
 Q92M78
 ID Q92M78 PRELIMINARY; PRT; 207 AA.
 AC Q92M78
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE Putative.
 GN JHP0344.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxID=85963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9912057; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori.";
 RL Nature 397:176-180(1999).
 DR EMBL; AE001470; AAD05923.1; -.
 KW Complete genome.
 SQ SEQUENCE 207 AA; 23527 MW; FFACCFEB7052624F CRC64;

Query Match 10.3%; Score 129; DB 16; Length 207;
 Best Local Similarity 25.0%; Pred. No. 0.16;
 Matches 53; Conservative 34; Mismatches 67; Indels 58; Gaps 10;

QY 42 LHFYIPKQKQPNKSHL-----VLIIEPKIIN-KVIPESYQKEFEKSLFLQSSFLERK 97
 DB 24 LHLKY-----KDYKPNSEKTKASTLTTPPKIFNAHFVPPFYQKEFKALAQCIAYFL----- 75

QY 98 YSVSQFQDASEIPQDIKEKALLVIRMDGNVAI-LEDIVESDALS-EKVIDMSSGYLN 154
 DB 76 -----KDKSALFNISGNVFFESPDKRAIKERLKTIEPNT----- 115

QY 155 LNFVEPKSDIHSFGIDVSKIKAVIER-----VELRRTNSSGFVPTTV 199
 DB 116 ----DPRA--VMRFLNQLASLILECVPTACPTDTLLIPTALSPIDYANRLGDNPSLFP 169

QY 200 HRIKETHDQAIKIMNOAYHKVAVHITKELS 231
 DB 170 Q--EDKSYHNALIKALNKAYISLMEGLEKRLN 199

RESULT 8
 Q9U459
 ID Q9U459 PRELIMINARY; PRT; 5458 AA.
 AC Q9U459; Q9GPG4;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Erythrocyte membrane-associated giant protein antigen 332.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FC01/HN;
 RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M., Lu J.H., X J.;
 RT "Molecular cloning and structure analysis of the Plasmodium falciparum
 RT erythrocyte membrane-associated giant protein Ag332 (Pf332) gene";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF202180; AAF15293.3; -.

DR InterPro; IPR006763; Ag332.
 DR InterPro; IPR001313; Pumilio/Puf.
 DR Pfam; PF04671; Ag332; 138.
 SQ SEQUENCE 5458 AA; 615269 MW; 4F6A37D92BD80172 CRC64;

Query Match 10.2%; Score 127.5; DB 5; Length 5458;
 Best Local Similarity 21.4%; Pred. No. 12;
 Matches 45; Conservative 47; Mismatches 55; Indels 63; Gaps 7;

QY 69 INKVIPESYQKEFEKSLFLQSSFLERKGYVSQFKDASEIPQDIKEKALLVLR----- 122
 DB 2568 IEEVVEEISTTEK-----LKEASAIEEFVEESIREDVLEESLVTVNVGQOE 2618

QY 123 -----MDGNVAILEDIVESDALSSEKVIDMSSGYLNLFVEPK-----S 162

DB 2619 SVTIEIVDGSFTEDIVEEESVTEIIVDEES--VTKEIVEDEELVTEIVEDEGSFT 2676

QY 163 EDIIEH---SFGIDVSKIKAVIERVELRTNSGGFVPKTFVHRIKETDHDQAIKIMNOAY 219
 DB 2677 BEIVEDEGSFTEEVIEERSLIEVEDTET-----VAEKEGSGVKEIIDE-- 2721

QY 220 HKVMWHITKEISKHMEHYKVSSEMKKRK 249
 DB 2722 -----KSUTEKIVEEESVTEIVEEKE 2743

RESULT 9
 QBTHN3
 ID Q8IHN3 PRELIMINARY; PRT; 5507 AA.
 AC Q8IHN3;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Antigen 332, putative.
 GN PF11_0507.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RX MEDLINE=22355705; PubMed=12368864;
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., K.,
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S.L., Paulsen I.T., James K.,
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
 RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
 RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
 RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
 RA Fraser C.W., Barrell B.;
 RT "Genome sequence of the human malaria parasite Plasmodium
 RT falciparum";
 RL Nature 419:498-511(2002).
 DR EMBL; AE014843; AAN36076.1; -.
 SQ SEQUENCE 5507 AA; 621156 MW; 03B095165D1490AE CRC64;

Query Match 10.2%; Score 127.5; DB 5; Length 5507;
 Best Local Similarity 21.4%; Pred. No. 12;
 Matches 45; Conservative 47; Mismatches 55; Indels 63; Gaps 7;

QY 69 INKVIPESYQKEFEKSLFLQSSFLERKGYVSQFKDASEIPQDIKEKALLVLR----- 122
 DB 2503 IEEVVEEISTTEK-----LKEASAIEEFVEESIREDVLEESLVTVNVGQOE 2553

QY 123 -----MDGNVAILEDIVESDALSSEKVIDMSSGYLNLFVEPK-----S 162

DB 2554 SVTIEIVDGSFTEDIVEEESVTEIIVDEES--VTKEIVEDEELVTEIVEDEGSFT 2611

QY 163 EDIIEH---SFGIDVSKIKAVIERVELRTNSGGFVPKTFVHRIKETDHDQAIKIMNOAY 219
 DB 2612 BEIVEDEGSFTEEVIEERSLIEVEDTET-----VAEKEGSGVKEIIDE-- 2656

QY 220 HKVWHITKELSKKHMEHYKVSSEMKRK 249
DB 2657 -----KSLTEKIVEEKSVTVEEKE 2678

RESULT 10
Q9NFS3
ID Q9NFS3 PRELIMINARY; PRT; 16215 AA.
AC Q9NFS3;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE D-titin.
GN SLU OR D-TITIN OR CG1915 OR CG18242 OR CG18245 OR CG18857.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang Y.Q., Broadie K.S.;
RT "Characterization of Drosophila D-Titin gene";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
ENBL; A021740; CAB9524.1; -;
DR HSSP; P56276; 1TLK.
DR FlyBase; FBgn0003432; sls.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IGc2.
DR InterPro; IPR003086; IG_VHC.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF00047; IG; 50.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00408; IGc2; 15.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00835; IG_LIKE; 50.
DR PROSITE; PS50002; SH3; 1.
KW Immunoglobulin domain; SH3 domain.
SQ SEQUENCE 16215 AA; 1841509 MW; 242C8765E00F7603 CRC64;

Query Match 9.3%; Score 116.5; DB 5; Length 16215;
Best Local Similarity 21.8%; Pred. No. 2.7e+02;
Matches 50; Conservative 48; Mismatches 56; Indels 77; Gaps 11;

QY 21 TALADGMPAKQOHNNTGESVELHFPYIKQKQEPKNSHLVLTIEPKIEINKVIPESYQKE 80
DB 12762 TATAQTTPGAQBEKST-----QDDTKDT-----IQKTVHKTKPDT-QKS 12801

QY 81 FEKSLFLOLSSFLERKYSVSQFQDASEIPQDIKEKALLVRLMDGNVALLD-----IVEBS 137
DB 12802 VETSELPEV-----HKDQYQISITHE--ELVEEQPEKILEVRVIDEVAEVSQPIVEEV 12854

QY 138 DALSEKVIDMSSGYLNLNFPVEPKSEDIHSFGIDYSKIKAVIERVELARTMSGGVPKPT 197
DB 12855 E--DEEP-----QATSEVTE-----DVTKEPKKKVKKTKTD----- 12896

QY 198 FVHRKETDHDQAIRKIMQAVHKVMVHTKELSKKHMEHYKVSSEMKRK 248
DB 12887 -----DHDELKIKMLEQ-----ETEKTELEKYEKIEFDVPEK 12918

RESULT 11
Q917U4
ID Q917U4 PRELIMINARY; PRT; 18074 AA.
AC Q917U4;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE CG1915 protein.
GN SLU OR CG1915 OR CG18242 OR CG18245 OR CG18857.

OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Berkelley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brotter P.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Cantor A., Chandra I.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Davies P.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.I., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegun C.,
RA Jalali M., Kalush C., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Kimball B.E., Kodira C., Kraft C., Kravitz S., Liang Y., Lin X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Mepherson D.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., Mepherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector A., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster";
RT Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banton J., An H., Baldwin D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Fartan D.,
RA Ferreria S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegun C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragov V., Park S., Patel S., Pfeiffer B., Scheeler F.,
RA Phuanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.B., Smith C.D.,
RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Celniker S.B.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;

RT "Annotation of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RA SEQUENCE FROM N.A.
 RP Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RA SEQUENCE FROM N.A.
 RP FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 DR EMBL; AE003473; AAG22226.2; --
 DR HSP; P56276; IILK.
 DR FlyBase; FBgn003432; sls.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR003599; IG.
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003598; IG C2.
 DR InterPro; IPR003006; IG_MHC.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00041; fn3; 5.
 DR Pfam; PF00047; ig; 39.
 DR Pfam; PF00018; SH3; 1.
 DR SMART; SM00409; IG; 53.
 DR SMART; SM00408; IGC2; 46.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS50835; IG_LIKE; 50.
 DR PROSITE; PS50002; SH3; 1.
 KW Immunoglobulin domain; SH3 domain.
 SQ SEQUENCE 18074 AA; 2058295 MW; 0710C589B3B9D050 CRC64;

Query Match 9.3%; Score 116.5; DB 5; Length 18074;
 Best Local Similarity 21.6%; Pred. No. 3e+02;
 Matches 50; Conservative 48; Mismatches 56; Indels 77; Gaps 11;
 YQ 21 TALADGMPAKQHNNTGESVELHFPYPIKGQPKNSHLVLIPIKINKEIPESYQKE 80
 DB 14621 TATAQTPTSAQDEKST-----QDDTKDT-----IQTKVKKTKPDT-QKS 14660
 YQ 81 FEKSLFLQSLFSLERKYSVSQFQKASEIPQDIKEKALLVLRMDGNVAILED---IYES 137
 DB 14661 VETSELPEV-----HKDYQISIIHE--ELVEESQPEKILSVRVIDEVAEESQPIVEEV 14713
 YQ 138 DALSEEVKIDMSGYLNLNFVEPKSEDIHISFGIDVSKIKAVIRVELRRTNSGGFVPKT 197
 DB 14714 E-DEEP-----QPATEETVE-----DVTKPSKKKKVKKKTD----- 14745
 YQ 198 FVHRKETDHDQAIRKIMNQAYHKVWHITKELSKHMEHYKVSSEMKKR 248
 DB 14746 -----DHDELIKKYLEQ-----EIEKTELEKYEIEFDPVPRK 14777

RESULT 12
 Q926X5 PRELIMINARY; PRT; 722 AA.
 ID Q926X5
 AC Q926X5
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein lin2915.
 GN LIN2915.
 OS Listeria innocua.
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 OX NCBI_TaxID=1642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLIP 11262 / Serovar 6a;
 RX PubMed=11679669;
 RA Glaser P., Frangoul L., Buchrieser C., Rueniok C., Amend A.,
 RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,

RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kutapkat G.,
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J. C., Purcell R.,
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
 RT "Comparative genomics of Listeria species.";
 RL Science 294:849-852(2001).
 DR EMBL; AL596174; CAC98140.1; --
 DR ListList; LIN02915; --
 DR InterPro; IPR006541; Bacteriocin_ass.
 DR TIGRfams; TIGR01854; bact_immun_7cm; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 722 AA; 83191 MW; ECC28EBDEE8A9811 CRC64;

Query Match 9.2%; Score 115.5; DB 16; Length 722;
 Best Local Similarity 28.4%; Pred. No. 6.4;
 Matches 63; Conservative 35; Mismatches 75; Indels 49; Gaps 14;
 YQ 37 GESVELHFPY-----PIKQKQPKNSHLVLIPIKINKEIPESYQKEFKSL-FLOLSS 91
 DB 427 GNSVEVDENFLKLNPIKSAQ--CONVENAINDKV-LNIIIVPTS-KKEFEKDIKKAYLDH 483
 YQ 92 FLERKYSVSQFQKASEIPQDIKEKALLVLRMDG--NVAILEDIVEESALSEKVIDM 148
 DB 484 FVFKQVEVANINNEALNP-----VLESKDLSINIYAQN---NQDYFS-----YDS 529
 YQ 149 SSGYLNL-NFVEPKSEDIHISFGIDVSKIKAVIRVELRRTNSGGFVPKTFVHRKETDH 207
 DB 530 SAGDLRTGNTIDFIA--LIYTGNDSSSIGA-----RVTSYVYVDKT-----KGDA 574
 YQ 208 DOAIRKIMNQAYHKVWHITKELSKHMEHYKVSSEMKKR 249
 DB 575 FNAIPLINNSNAREITNVTSV-----YQEVSSSELTALK 608

RESULT 13
 Q9NWC6 PRELIMINARY; PRT; 535 AA.
 ID Q9NWC6
 AC Q9NWC6
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE Hypothetical protein (fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Negai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
 RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
 RA Ninomiya K., Iwayanagi T.;
 RT "NEDO human cDNA sequencing project";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK000990; BA91457.1; --
 FT NON TER 535
 SQ SEQUENCE 535 AA; 62604 MW; C50CF36B52791101 CRC64;

Query Match 9.1%; Score 113.5; DB 4; Length 535;
 Best Local Similarity 22.7%; Pred. No. 6.1;
 Matches 66; Conservative 43; Mismatches 87; Indels 95; Gaps 14;
 YQ 30 KQHNNTGESVELHFPY-PIKQKQPKNSHLVLIPIKINKEIPESYQKEFKSLFLO 88
 DB 79 EQAHNLT---IEMKNHYVPLKVSSEDMKSHDAII---DDLNRKLLDVTQKTEKLEME 131
 YQ 89 LSSFLERKYSVSQFQKASE-----IQDIKEKALLVLRMDGNVAILEDIVERDALSE- 142

```
Db 132 -KLLLENDLSL-----KQSRLETVPPEKEHEIALK-----SNIVELKQLSEL 178
QY 143 -----EKVIDMSSGYLNIN-----FVEPKS 162
Db 179 KKKCGEDQEKIHALTSNTNLKQMSNQYVPVKTHEEVKMTLNDTLAKTNRELLDVKKKF 238
QY 163 EDIHSFGIDVSKIKAVIERVELRRTNSGGFPVKTFV-----HRIKETDHDQAIRKIMNQ- 217
Db 239 EDINQEF-----VKIKDKNEILKRNLENTQNIKAEYISLAEHEAKMSSLSQSMRKVQDSN 294
QY 218 -----AYHK-----VMVHTKLSKKHMEHYEK-----VSSEMKKRK 249
Db 295 AEILANYRKQOEIVTLHAEIKAKQKELDITQECIKVKYAPIVSFECECRK 345

RESULT 14
Q9HCL1 PRELIMINARY; PRT; 1416 AA.
AC Q9HCL1 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 23, Last annotation update)
DE KIAA1561 protein (Fragment).
GN KIAA1561.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20450693; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirose M., Chara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVIII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.";
RL DNA Res. 7:273-281(2000).
DR EMBL; AB046781; BAB13387.2; -.
DR HSSP; P04268; 1IC2.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000727; T_SNARE.
DR Pfam; PF00023; ank; 5; T_SNARE.
DR PRINTS; PR01415; ANKYRN.
DR SMART; SM00248; ANK; 6.
DR PROSITE; PS50088; ANK REPEAT; 5.
DR PROSITE; PS50297; ANK REP REGION; 1.
DR PROSITE; PS50192; T_SNARE; 1.
KW ANK repeat; Repeat.
FT NON TER 1
SQ SEQUENCE 1416 AA; 162505 MW; 92C1104445FE0009 CRC64;

Query Match 9.1%; Score 113.5; DB 4; Length 1416;
Best Local Similarity 22.7%; Pred. No. 21;
Matches 56; Conservative 43; Mismatches 87; Indels 95; Gaps 14;

QY 30 KQHNNTGESVELHFHY-PIKQKQEPKNSHLVLIPEKIEINKVIPESYQKEFKSLFLQ 88
Db 731 EQAHNLT---IEMKNHYVPLKVSDEMKSHDAII---DDLNRKLLDVTKYTEKKLEME 783
QY 89 LSSFLERKGYVSQFKDASE-----IPQDIKEKALLVLRMDGNVAILEDIVESDALSE- 142
Db 784 -KLLLENDLSL-----KQSRLETVPPEKEHEIALK-----SNIVELKQLSEL 830
QY 143 -----EKVIDMSSGYLNIN-----FVEPKS 162
Db 831 KKKCGEDQEKIHALTSNTNLKQMSNQYVPVKTHEEVKMTLNDTLAKTNRELLDVKKKF 890
QY 163 EDIHSFGIDVSKIKAVIERVELRRTNSGGFPVKTFV-----HRIKETDHDQAIRKIMNQ- 217
Db 891 EDINQEF-----VKIKDKNEILKRNLENTQNIKAEYISLAEHEAKMSSLSQSMRKVQDSN 946
QY 218 -----AYHK-----VMVHTKLSKKHMEHYEK-----VSSEMKKRK 249
Db 295 AEILANYRKQOEIVTLHAEIKAKQKELDITQECIKVKYAPIVSFECECRK 997

Search completed: December 10, 2003, 18:36:35
Job time : 43.1767 sec
```

```
Db 947 AEILANYRKQOEIVTLHAEIKAKQKELDITQECIKVKYAPIVSFECECRK 997

RESULT 15
Q9BZF9 PRELIMINARY; PRT; 1416 AA.
AC Q9BZF9 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE Uveal autoantigen.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21092738; PubMed=11162650;
RA Yamada K., Serju S., Nakatsura T., Murata Y., Ishihara M.,
RA Nakamura S., Ohno S., Negi A., Nishimura Y.;
RT "Identification of a novel autoantigen UACA in patients with
RT panuveitis.";
RL Biochem. Biophys. Res. Commun. 280:1169-1176(2001).
DR EMBL; AF322916; AAG49577.1; -.
DR HSSP; Q00420; 1AWC.
DR Genew; HGNC:15947; UACA.
DR InterPro; IPR02110; ANK.
DR InterPro; IPR000727; T_SNARE.
DR Pfam; PF00023; ank; 6; T_SNARE.
DR SMART; SM00248; ANK; 6.
DR PROSITE; PS50088; ANK REPEAT; 5.
DR PROSITE; PS50297; ANK REP REGION; 1.
DR PROSITE; PS50192; T_SNARE; 1.
KW ANK repeat; Repeat.
SQ SEQUENCE 1416 AA; 162579 MW; DAE2F5B54AA76BD CRC64;

Query Match 9.1%; Score 113.5; DB 4; Length 1416;
Best Local Similarity 22.7%; Pred. No. 21;
Matches 56; Conservative 43; Mismatches 87; Indels 95; Gaps 14;

QY 30 KQHNNTGESVELHFHY-PIKQKQEPKNSHLVLIPEKIEINKVIPESYQKEFKSLFLQ 88
Db 731 EQAHNLT---IEMKNHYVPLKVSDEMKSHDAII---DDLNRKLLDVTKYTEKKLEME 783
QY 89 LSSFLERKGYVSQFKDASE-----IPQDIKEKALLVLRMDGNVAILEDIVESDALSE- 142
Db 784 -KLLLENDLSL-----KQSRLETVPPEKEHEIALK-----SNIVELKQLSEL 830
QY 143 -----EKVIDMSSGYLNIN-----FVEPKS 162
Db 831 KKKCGEDQEKIHALTSNTNLKQMSNQYVPVKTHEEVKMTLNDTLAKTNRELLDVKKKF 890
QY 163 EDIHSFGIDVSKIKAVIERVELRRTNSGGFPVKTFV-----HRIKETDHDQAIRKIMNQ- 217
Db 891 EDINQEF-----VKIKDKNEILKRNLENTQNIKAEYISLAEHEAKMSSLSQSMRKVQDSN 946
QY 218 -----AYHK-----VMVHTKLSKKHMEHYEK-----VSSEMKKRK 249
Db 947 AEILANYRKQOEIVTLHAEIKAKQKELDITQECIKVKYAPIVSFECECRK 997

Search completed: December 10, 2003, 18:36:35
Job time : 43.1767 sec
```


GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:28:11 ; Search time 22.3325 Seconds
(without alignments)
878.747 Million cell updates/sec

Title: US-10-080-113-2

Perfect score: 906

Sequence: 1 MKRSSVFSLVAFLLVAGCS.....AQKTRCYKRNRRVDVLMK 179

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- A_Geneseq_19Jun03.*
- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
 - 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
 - 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
 - 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
 - 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
 - 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
 - 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
 - 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
 - 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
 - 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
 - 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
 - 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
 - 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
 - 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
 - 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
 - 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
 - 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
 - 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
 - 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
 - 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
 - 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
 - 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
 - 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
 - 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	906	100.0	179	20 AAW89982	Expressed antigen
2	906	100.0	179	22 AAB46316	H. pylori HPS144 p
3	891	98.3	179	20 AAW89813	Protein encoded by
4	877	96.8	179	18 AAW20394	H. pylori outer me
5	877	96.8	179	18 AAW24651	H. pylori outer me
6	877	96.8	187	18 AAW20795	H. pylori outer me
7	836	92.3	179	18 AAW23592	H. pylori Cmp22 ou
8	691	76.3	144	18 AAW23591	H. pylori recombin
9	532	58.7	104	20 AAW89983	Expressed antigen

10	221.5	24.4	172	21	AAV55090	M. catarrhalis BAS
11	220.5	24.3	172	21	AAV55091	M. catarrhalis BAS
12	218.5	24.1	172	21	AAV55092	M. catarrhalis BAS
13	215.5	23.8	172	21	AAV55089	M. catarrhalis BAS
14	215.5	23.8	326	22	ABBS9178	C.celi CadF protei
15	208	23.0	242	23	ABG80415	Neisseria meningit
16	194.5	21.5	236	23	ABG91067	Neisseria gonorrhoe
17	194.5	21.5	236	24	ABP80776	N. gonorrhoeae ami
18	183	20.2	153	9	AP82947	16600 dalton outer
19	183	20.2	153	9	AP80593	16600 dalton outer
20	183	20.2	153	11	AAK05797	PBCMP-1 gene produ
21	183	20.2	153	15	AAK51161	Outer membrane pro
22	183	20.2	153	23	ABG80421	Haemophilus influe
23	183	20.2	153	23	ABG80422	Haemophilus influe
24	180.5	19.9	287	22	ABBS9182	C.jejuni cadF-F380
25	179.5	19.8	153	11	ABR07145	16 kD outer membr
26	179.5	19.8	287	22	ABBS9181	C.jejuni cadF-M129
27	178	19.6	153	9	AP80665	Sequence encoded H
28	169	18.7	335	16	AAK88257	K.pneumoniae immun
29	169	18.7	335	17	AAK95644	Klebsiella pneumon
30	169	18.7	335	18	AAK93796	Protein P40, and O
31	169	18.7	344	18	AAK93797	Protein LP40, a va
32	169	18.7	344	20	AAK44077	K.pneumoniae OmpA
33	169	18.7	344	21	AAK18934	A P40 polypeptide
34	169	18.7	344	21	AAK18904	A Klebsiella pneum
35	169	18.7	344	21	AAK08825	A P40 polypeptide
36	169	18.7	344	21	AAK08317	An outer membrane
37	169	18.7	344	21	AAK08341	An outer membrane
38	169	18.7	344	21	AAK93341	Amino acid sequenc
39	169	18.7	344	22	AAK67743	Amino acid sequenc
40	169	18.7	344	22	AAK63698	Amino acid sequenc
41	169	18.7	344	22	AAK84122	Amino acid sequenc
42	169	18.7	344	22	AAK67770	Amino acid sequenc
43	169	18.7	344	23	AAK48395	Klebsiella pneumon
44	169	18.7	344	23	AAK47796	Klebsiella pneumon
45	169	18.7	452	22	AAK67771	Amino acid sequenc

ALIGNMENTS

RESULT 1

AAW89982

ID AAW89982 standard; Protein: 179 AA.

XX AAW89982;

DT 20-MAR-2003 (updated)

DT 18-FEB-1999 (first entry)

XX

DE Expressed antigen for clone Y175A.

XX

KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
peptic ulcer; gastric adenocarcinoma; gastric lymphoma.

XX Helicobacter pylori.

XX WO9849314-A2.

XX 05-NOV-1998.

XX 25-APR-1998; 98WO-US08487.

XX 14-OCT-1997; 97US-0061958.

XX 25-APR-1997; 97US-0045107.

(GENE-) GENELABS TECHNOLOGIES INC.

Chow TP, Fry KE, Lim MY, McAtee CP;

WPI; 1999-009433/01.

New Helicobacter pylori antigens and related nucleic acid sequences

PT - useful in serological diagnosis and protective vaccines, providing
 PT long-lasting immune response

XX Claim 16; Page 321; 402pp; English.

XX The present sequence represents a Helicobacter pylori antigenic protein
 CC that is characterised by immunoreactivity with H. pylori-positive
 CC antisera. The proteins are highly immunogenic and induce a long-lasting
 CC immune response that persists even after antimicrobial treatment. In
 CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
 CC highly sensitive and specific. The specification also describes 69
 CC previously unrecognised immunogenic cluster families. H. pylori antigens
 CC are used to detect H. pylori-specific antibodies, for diagnosing
 CC infection or to confirm eradication of infection, and in vaccines to
 CC protect against H. pylori infection and related diseases (gastritis,
 CC peptic ulcer, gastric adenocarcinoma/lymphoma).
 CC (Updated on 20-MAR-2003 to correct PF field.)

XX Sequence 179 AA;

Query Match 100.0%; Score 906; DB 20; Length 179;
 Best Local Similarity 100.0%; Pred. No. 9.3e-81;
 Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRSSVFLVAFLLVAGCSHKMDNKTVDGVSATVQTAPVTTEPAPEKEEPKQEPAPV 60
 DB 1 MKRSSVFLVAFLLVAGCSHKMDNKTVDGVSATVQTAPVTTEPAPEKEEPKQEPAPV 60
 QY 61 VEKPAVESGTHIIASIVDFDKYIEKESDQETLDEIVQKAKENHMQVLLGNTDFGSSE 120
 DB 61 VEKPAVESGTHIIASIVDFDKYIEKESDQETLDEIVQKAKENHMQVLLGNTDFGSSE 120
 QY 121 YNQALGVKRTLSVKNALVIGVCKMDIKTISFGETPKCAQKTRCYKRNRRVDVKLMK 179
 DB 121 YNQALGVKRTLSVKNALVIGVCKMDIKTISFGETPKCAQKTRCYKRNRRVDVKLMK 179

RESULT 2

AAAB46316
 ID AAB46316 standard; Protein; 179 AA.

XX AAB46316;

XX 05-APR-2001 (first entry)

XX H. pylori HPS144 protein.

XX Microbial infection; antibacterial; Helicobacter pylori infection;
 KW vaccine; screening.

XX Helicobacter pylori.

XX WO200073502-A2.

XX 07-DEC-2000.

XX 31-MAY-2000; 2000WO-BP05024.

XX 31-MAY-1999; 99DE-1024965.

XX 17-JUN-1999; 99DE-1027740.

XX 21-JUL-1999; 99DE-1034029.

XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA (CREA-) CREATOGEN GMBH.

XX Apfel H, Fuchs TM, Gibbs CF, Hueck CJ, Meyer TF;

XX WPI; 2001-049948/06.

XX N-PSDB; AAF25593.

XX Preparing an agent for diagnosis or control of microbial infection,
 PT useful particularly against Helicobacter, based on identification of
 PT essential genes in defective mutants

XX Claim 37; Page 253; 366pp; German.

XX This invention describes a novel preparation of an agent (A) for
 CC detection, prevention and/or treatment of microbial infection by:
 CC (i) identifying essential genes (I) and corresponding polypeptides
 CC (II); (ii) identifying compounds that are directed against (II) and
 CC inactivate the microbe; (iii) testing these for suitability for use; and
 CC (iv) formulating selected (A). Identifying essential genes (I) comprises
 CC preparation of gene-deficient microorganisms by conditional antisense
 CC inhibition (CAI) and/or subtractive recombination mutagenesis (SRM).
 CC Then determining viability and/or survival of the deficient organisms.
 CC The products of the invention have antibacterial activity. (A) which may
 CC be a nucleic acid (Ia), vector or host cell containing (Ia), derived
 CC polypeptide (IIa), or fragments, (IIa)-specific antibodies or their
 CC fragments or an inhibitor of (IIa) are particularly used for diagnosis,
 CC treatment or prevention of infection by Helicobacter pylori. Particularly
 CC (Ia) and (IIa) are used in DNA, subunit or live vaccines. The method
 CC identifies essential genes, including those that have homologs in other
 CC species, so identified (A) should have a broad spectrum of activity. Many
 CC gene-deficient cells can be screened quickly, in an automated process,
 CC and the identified genes can be used for screening without purification.

XX Sequence 179 AA;

Query Match 100.0%; Score 906; DB 22; Length 179;
 Best Local Similarity 100.0%; Pred. No. 9.3e-81;
 Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRSSVFLVAFLLVAGCSHKMDNKTVDGVSATVQTAPVTTEPAPEKEEPKQEPAPV 60

DB 1 MKRSSVFLVAFLLVAGCSHKMDNKTVDGVSATVQTAPVTTEPAPEKEEPKQEPAPV 60

QY 61 VEKPAVESGTHIIASIVDFDKYIEKESDQETLDEIVQKAKENHMQVLLGNTDFGSSE 120

DB 61 VEKPAVESGTHIIASIVDFDKYIEKESDQETLDEIVQKAKENHMQVLLGNTDFGSSE 120

QY 121 YNQALGVKRTLSVKNALVIGVCKMDIKTISFGETPKCAQKTRCYKRNRRVDVKLMK 179

DB 121 YNQALGVKRTLSVKNALVIGVCKMDIKTISFGETPKCAQKTRCYKRNRRVDVKLMK 179

RESULT 3

AAW89813

ID AAW89813 standard; Protein; 179 AA.

XX AAW89813;

XX 20-MAR-2003 (updated)

DT 18-FEB-1999 (first entry)

XX Protein encoded by clone G1a ORF2.

XX Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 KW peptic ulcer; Gastric adenocarcinoma; Gastric lymphoma.

XX Helicobacter pylori.

XX WO9849314-A2.

XX 05-NOV-1998.

XX 25-APR-1998; 98WO-US08487.

XX 14-OCT-1997; 97US-0061958.

XX 25-APR-1997; 97US-0045107.

XX (GENE-) GENELABS TECHNOLOGIES INC.

XX Chow TP, Fry KE, Lim MY, McAtee CP;

XX WPI; 1999-009433/01.

XX N-PSDB; AAV90545.

XX New Helicobacter pylori antigens and related nucleic acid sequences
PT - useful in serological diagnosis and protective vaccines, providing
XX long-lasting immune response
PS Claim 10; Page 96-97; 402pp; English.
XX
XX The present sequence represents a Helicobacter pylori antigenic protein
CC that is characterised by immunoreactivity with H. pylori-positive
CC antisera. The proteins are highly immunogenic and induce a long-lasting
CC immune response that persists even after antimicrobial treatment. In
CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC highly sensitive and specific. The specification also describes 69
CC previously unrecognised immunogenic cluster families. H. pylori antigens
CC are used to detect H. pylori-specific antibodies, for diagnosing
CC infection or to confirm eradication of infection, and in vaccines to
CC protect against H. pylori infection and related diseases (gastritis,
CC peptic ulcer, gastric adenocarcinoma/lymphoma).
CC (Updated on 20-MAR-2003 to correct PF field.)
XX
XX Sequence 179 AA;
SQ
Query Match 98.3%; Score 891; DB 20; Length 179;
Best Local Similarity 97.8%; Pred. No. 2.8e-79;
Matches 175; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 MKRSSVPSFLVAFLLVAGCSHRMDNKTVDGVSADTQVTPVTTTTPAPEKEEPKQEPAPV 60
DB 1 MKRSSAFSLVAFLLVAGCSHRMDNKTVDGVSADTQVTPVTTTTPAPEKEEPKQEPAPV 60
QY 61 VEEKPAVESGTTIIASIFDFDKYIKESDQETLDEIVQAKENHMQVLLGNTDFGSGS 120
DB 61 VEEKPAIESGTTIIASIFDFDKYIKESDQETLDEIVQAKENHMQVLLGNTDFGSGS 120
QY 121 YNQALGVKRTLSVKNALVTKGVEKDMIKTISFGETKPKCAQKTRCYKRNRRVDVKLMK 179
DB 121 YNQALGVKRTLSVKNALVTKGVEKDMIKTISFGETKPKCAQKTRCYKRNRRVDVKLVK 179
RESULT 4
AAW20394
ID AAW20394 standard; Protein; 179 AA.
XX
AC AAW20394;
XX
XX 29-JUL-1997 (first entry)
DT
DE H. pylori outer membrane protein 31262.aa.
XX
KW Outer membrane; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacterium; life cycle; activator;
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW diagnosis.
XX
OS Helicobacter pylori.
XX
PN WO9640893-A1.
XX
XX 19-DEC-1996.
PD
XX
XX 06-JUN-1996; 96WO-US09122.
PF
XX
XX 01-APR-1996; 96US-0630405.
PR
XX 07-JUN-1995; 95US-0487032.
PR
XX
XX (ASTR) ASTRA AB.
PA
XX
XX Berglindh OT, Smith D, Mellgaard BL;
PI
XX
XX WPI; 1997-052306/05.
DR
XX N-PSDB; AAT67789.
DR
XX Helicobacter pylori nucleic acid sequences and related
PT

PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
XX
PS Claim 56; Page 577; 1481pp; English.
XX
XX This sequence is a H. pylori outer membrane protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
XX
XX Sequence 179 AA;
SQ
Query Match 96.8%; Score 877; DB 18; Length 179;
Best Local Similarity 95.5%; Pred. No. 6.6e-78;
Matches 171; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 1 MKRSSVPSFLVAFLLVAGCSHRMDNKTVDGVSADTQVTPVTTTTPAPEKEEPKQEPAPV 60
DB 1 MKRSSAFSLVAFLLVAGCSHRMDNKTVDGVSADTQVTPVTTTTPAPEKEEPKQEPAPV 60
QY 61 VEEKPAVESGTTIIASIFDFDKYIKESDQETLDEIVQAKENHMQVLLGNTDFGSGS 120
DB 61 VEEKPAIESGTTIIASIFDFDKYIKESDQETLDEIVQAKENHMQVLLGNTDFGSGS 120
QY 121 YNQALGVKRTLSVKNALVTKGVEKDMIKTISFGETKPKCAQKTRCYKRNRRVDVKLMK 179
DB 121 YNQALGVKRTLSVKNALVTKGVEKDMIKTISFGESKPKCVQKTRCYKRNRRVDVKLVK 179
RESULT 5
AAW24651
ID AAW24651 standard; Protein; 179 AA.
XX
AC AAW24651;
XX
XX 11-AUG-1997 (first entry)
DT
DE H. pylori outer membrane protein 31262.aa.
XX
KW Transmembrane; cytoplasmic; cell envelope; flagella; transport;
KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
KW activator; inhibitor; bacterial life cycle; vaccine; immunise;
KW detection; antisense; inhibition.
XX
OS Helicobacter pylori.
XX
PN WO9719098-A1.
XX
XX 29-MAY-1997.
PD
XX
XX 15-NOV-1996; 96WO-US18542.
PF
XX
XX 17-NOV-1995; 95US-0561469.
PR
XX
XX (ASTR) ASTRA AB.
PA
XX
XX Smith DH;
PI
XX
XX WPI; 1997-298052/27.
DR
XX N-PSDB; AAT77469.
DR
XX Helicobacter pylori nucleic acid sequences and related proteins -
PT used for diagnostics and therapeutics

XX PS Claim 18; Page 171; 235pp; English.
 XX CC This sequence represents an H. pylori outer membrane protein.
 CC CC Helicobacter pylori has been strongly linked to chronic gastritis and
 CC CC duodenal ulcer disease. The nucleic acid sequences of the invention
 CC CC are used to evaluate compounds, especially activators or inhibitors of
 CC CC bacterial life cycle, for the ability to bind an H. pylori nucleic acid
 CC CC sequence. The nucleic acid sequences, and corresponding proteins, are
 CC CC also useful for generating vaccines for immunising subjects against H.
 CC CC pylori or for use in detecting the presence of Helicobacter species in
 CC CC a sample. Antisense nucleic acid sequences of these sequences are
 CC CC used to inhibit expression of a gene from Helicobacter species. H.
 CC CC pylori whole genomic DNA was isolated and nebulised to a median size of
 CC CC 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique
 CC CC BstXI-linker adapters in 100-1000 fold molar excess. These linkers are
 CC CC complementary to the BstXI-cut pVPX vectors, while the overhang is not
 CC CC self-complementary. Therefore the linkers will not concatenate nor
 CC CC will the cut vector re-ligate itself easily. The linker-adaptor inserts
 CC CC were ligated to each of the 20 pMPX vectors to construct a series of
 CC CC shotgun subclone libraries. The purified DNA samples were then
 CC CC sequenced.
 CC CC Note: The ORF/protein reference number for this sequence was obtained
 CC CC from the related specification, WO9640893.
 XX SQ Sequence 179 AA;

Query Match 96.8%; Score 877; DB 18; Length 179;
 Best Local Similarity 95.5%; Pred. No. 6.6e-78;
 Matches 171; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MKRSSVFSFLVAFLLVAGCSHKMDNKTVAGDVSAKTQTAPVTTEPAPEKEEPKQEPAPV 60
 DB 1 MKRSSVFSFLVAFLLVAGCSHKMDNKTVAGDVSAKTQTAPVTTEPAPEKEEPKQEPAPV 60
 QY 61 VEEKPAVESGTTIIASIVDFDKYEIKESDQETLDEIVQKAKENHMVLLGNTDFGSGS 120
 DB 61 VEEKPAVESGTTIIASIVDFDKYEIKESDQETLDEIVQKAKENHMVLLGNTDFGSGS 120
 QY 121 YNOALGVKRTLSVKNALVIKGVKMDIKTISFGETKPKCAQKTRCYKENERVDVKLMK 179
 DB 121 YNOALGVKRTLSVKNALVIKGVKMDIKTISFGESKPKCVQKTRCYRNRVDVKLVK 179

RESULT 6
 AAW20795
 ID AAW20795 standard; Protein; 187 AA.
 XX AC AAW20795;
 XX DT 21-JUL-1997 (first entry)
 XX DE H. pylori outer membrane protein 07gp31516orf4.
 XX KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 OS Helicobacter pylori.
 XX PH Key Location/Qualifiers
 FT Peptide 1..35
 FT Protein 36..179
 FT /label= leader_sequence
 FT /label= mature_peptide
 FT /note= "Omp22"
 XX PN WO9640893-A1.
 XX PD 19-DEC-1996.
 XX PF 06-JUN-1996; 96WO-US09122.
 XX PR 01-APR-1996; 96US-0630405.
 XX PR 07-JUN-1995; 95US-0487032.
 XX FA (ASTR) ASTRA AB.
 XX PI Berglindh OT, Smith D, Melligaard BL;
 XX

DR WPI; 1997-052306/05.
 DR N-PSDB; AAT68048.
 XX PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX PS Claim 56; Page 1203; 1481pp; English.
 XX CC This sequence is a H. pylori outer membrane protein.
 CC CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC CC infection or to identify H. pylori polypeptide binding compounds,
 CC CC useful as potential H. pylori life cycle activators or inhibitors.
 CC CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC CC overlapping contigs generated by mechanically shearing the bacterial
 CC CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC CC and the predicted coding regions defined by computer evaluation. To
 CC CC identify likely H. pylori antigens for vaccine development, the amino
 CC CC acid sequences predicted from various ORF were analysed for significant
 CC CC homology to other known or exported membrane proteins. Having identified
 CC CC and determined the sequences of interest, particular regions can be
 CC CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC CC production, e.g. in E. coli hosts.
 XX SQ Sequence 187 AA;

Query Match 96.8%; Score 877; DB 18; Length 187;
 Best Local Similarity 95.5%; Pred. No. 7e-78;
 Matches 171; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MKRSSVFSFLVAFLLVAGCSHKMDNKTVAGDVSAKTQTAPVTTEPAPEKEEPKQEPAPV 60
 DB 9 MKRSSVFSFLVAFLLVAGCSHKMDNKTVAGDVSAKTQTAPVTTEPAPEKEEPKQEPAPV 68
 QY 61 VEEKPAVESGTTIIASIVDFDKYEIKESDQETLDEIVQKAKENHMVLLGNTDFGSGS 120
 DB 69 VEEKPAVESGTTIIASIVDFDKYEIKESDQETLDEIVQKAKENHMVLLGNTDFGSGS 128
 QY 121 YNOALGVKRTLSVKNALVIKGVKMDIKTISFGETKPKCAQKTRCYKENERVDVKLMK 179
 DB 129 YNOALGVKRTLSVKNALVIKGVKMDIKTISFGESKPKCVQKTRCYRNRVDVKLVK 187

RESULT 7
 AAW23592
 ID AAW23592 standard; Protein; 179 AA.
 XX AC AAW23592;
 XX DT 23-JAN-1998 (first entry)
 XX DE H. pylori Omp22 outer membrane protein.
 XX KW Omp22 gene; outer membrane protein; antigen; immunogen; stomach;
 KW gastritis; peptic ulcers; gastric cancer; diagnostic; vaccine.
 XX OS Helicobacter pylori.
 XX PH Key Location/Qualifiers
 FT Peptide 1..35
 FT Protein 36..179
 FT /label= leader_sequence
 FT /label= mature_peptide
 FT /note= "Omp22"
 XX PN WO9728264-A1.
 XX PD 07-AUG-1997.
 XX PF 06-SEP-1996; 96WO-KR00154.
 XX PR 30-JAN-1996; 96KR-0002105.
 XX

PA (MOGA-) MOGAM BIOTECHNOLOGY RES INST.
 XX Kim J, Moon H, Park Y, Seo W, Yu G, Yum J;
 XX WPI; 1997-402617/37.
 DR N-PSDB; AAT74195.
 XX Helicobacter pylori outer membrane protein, Omp22 - used in a
 PT vaccine for treatment or prevention of H. pylori infection
 XX
 PS Example 3; Fig 5; 38pp; English.
 CC This sequence represents Omp22, an outer membrane protein which
 CC exhibits antigenicity and immunogenicity against Helicobacter pylori.
 CC This bacterium is associated with inflammation of the stomach and
 CC gastritis related diseases e.g. peptic ulcers and gastric cancer. Omp22
 CC could be used in immunological therapy as a H. Pylori-specific antigen
 CC for the treatment and prevention of diseases associated with this
 CC microorganism e.g. as the active ingredient in a diagnostic kit or
 CC vaccine.
 XX
 SQ Sequence 179 AA;
 Query Match 92.3%; Score 836; DB 18; Length 179;
 Best Local Similarity 91.6%; Pred. No. 6.9e-74;
 Matches 164; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
 QY 1 MKRSSVFSLVAFLLVAGCSHKMDNKTIVAGDYSAKTVQTPVTTTPAPEKEPKQEPAPV 60
 DB 1 MKRSSVFGFLVAFLLVAGCSHKMDNKTIVAGDYSAKAVQSPVSTIAQEKQEPAPV 60
 QY 61 VEEKPAVESGTTIIASIFDFDKYEIKESQDQLDEIVQAKENHMQVLLGNTDFGSGSE 120
 DB 61 VEEKPAVESGTTIIASIFDFDKYEIKESQDQLDEIVQAKENHMQVLLGNTDFGSGSE 120
 QY 121 YNQALGVKRTLSVKNALVKGVEKMDIKTISFGETKPKCAQKTRCYKRNRRVDVVKLMK 179
 DB 121 YNQALGVKRTLSVKNALVKGVEKMDIKTISFGETKPKCTQKTRCYKRNRRVDVVKLVK 179
 RESULT 8
 AAW23591
 ID AAW23591 standard; Protein; 144 AA.
 AC AAW23591;
 XX 23-JAN-1998 (first entry)
 DE H. pylori recombinant Omp22 recombinant outer membrane protein.
 XX
 KW Omp22 gene; outer membrane protein; antigen; immunogen; stomach;
 KW gastritis; peptic ulcers; gastric cancer; diagnostic; vaccine;
 KW recombinant protein.
 XX
 OS Helicobacter pylori.
 XX
 FH Key Location/Qualifiers
 FT Protein 1..144
 FT /label= Omp22
 FT /note= "Recombinant outer membrane protein"
 XX
 PN KO9728264-A1.
 XX
 PD 07-AUG-1997.
 XX
 PF 06-SEP-1996; 96WO-KR00154.
 XX
 PR 30-JAN-1996; 96KR-0002105.
 XX
 PA (MOGA-) MOGAM BIOTECHNOLOGY RES INST.
 XX Kim J, Moon H, Park Y, Seo W, Yu G, Yum J;
 XX WPI; 1997-402617/37.
 DR N-PSDB; AAT74195.
 XX Helicobacter pylori outer membrane protein, Omp22 - used in a
 PT vaccine for treatment or prevention of H. pylori infection
 XX
 PS Example 3; Fig 5; 38pp; English.
 CC This sequence represents Omp22, an outer membrane protein which
 CC exhibits antigenicity and immunogenicity against Helicobacter pylori.
 CC This bacterium is associated with inflammation of the stomach and
 CC gastritis related diseases e.g. peptic ulcers and gastric cancer. Omp22
 CC could be used in immunological therapy as a H. Pylori-specific antigen
 CC for the treatment and prevention of diseases associated with this
 CC microorganism e.g. as the active ingredient in a diagnostic kit or
 CC vaccine.
 XX
 SQ Sequence 179 AA;
 Query Match 92.3%; Score 836; DB 18; Length 179;
 Best Local Similarity 91.6%; Pred. No. 6.9e-74;
 Matches 164; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
 QY 1 MKRSSVFSLVAFLLVAGCSHKMDNKTIVAGDYSAKTVQTPVTTTPAPEKEPKQEPAPV 60
 DB 1 MKRSSVFGFLVAFLLVAGCSHKMDNKTIVAGDYSAKAVQSPVSTIAQEKQEPAPV 60
 QY 61 VEEKPAVESGTTIIASIFDFDKYEIKESQDQLDEIVQAKENHMQVLLGNTDFGSGSE 120
 DB 61 VEEKPAVESGTTIIASIFDFDKYEIKESQDQLDEIVQAKENHMQVLLGNTDFGSGSE 120
 QY 121 YNQALGVKRTLSVKNALVKGVEKMDIKTISFGETKPKCAQKTRCYKRNRRVDVVKLMK 179
 DB 121 YNQALGVKRTLSVKNALVKGVEKMDIKTISFGETKPKCTQKTRCYKRNRRVDVVKLVK 179
 RESULT 8
 AAW23591
 ID AAW23591 standard; Protein; 144 AA.
 AC AAW23591;
 XX 23-JAN-1998 (first entry)
 DE H. pylori recombinant Omp22 recombinant outer membrane protein.
 XX
 KW Omp22 gene; outer membrane protein; antigen; immunogen; stomach;
 KW gastritis; peptic ulcers; gastric cancer; diagnostic; vaccine;
 KW recombinant protein.
 XX
 OS Helicobacter pylori.
 XX
 FH Key Location/Qualifiers
 FT Protein 1..144
 FT /label= Omp22
 FT /note= "Recombinant outer membrane protein"
 XX
 PN KO9728264-A1.
 XX
 PD 07-AUG-1997.
 XX
 PF 06-SEP-1996; 96WO-KR00154.
 XX
 PR 30-JAN-1996; 96KR-0002105.
 XX
 PA (MOGA-) MOGAM BIOTECHNOLOGY RES INST.
 XX Kim J, Moon H, Park Y, Seo W, Yu G, Yum J;
 XX WPI; 1997-402617/37.
 DR N-PSDB; AAT74195.
 XX Helicobacter pylori outer membrane protein, Omp22 - used in a
 PT vaccine for treatment or prevention of H. pylori infection
 XX
 PS Example 3; Fig 5; 38pp; English.
 CC This sequence represents Omp22, an outer membrane protein which
 CC exhibits antigenicity and immunogenicity against Helicobacter pylori.
 CC This bacterium is associated with inflammation of the stomach and
 CC gastritis related diseases e.g. peptic ulcers and gastric cancer. Omp22
 CC could be used in immunological therapy as a H. Pylori-specific antigen
 CC for the treatment and prevention of diseases associated with this
 CC microorganism e.g. as the active ingredient in a diagnostic kit or
 CC vaccine.
 XX
 SQ Sequence 144 AA;
 Query Match 76.3%; Score 691; DB 18; Length 144;
 Best Local Similarity 95.1%; Pred. No. 8.6e-60;
 Matches 136; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 QY 37 VQTAPVTEPAPEKEPKQEPAPVVEEKPAVESGTTIIASIFDFDKYEIKESQDQLDEI 96
 DB 2 VQSPVSTIAQEKQEPAPVVEEKPAVESGTTIIASIFDFDKYEIKESQDQLDEI 61
 QY 97 VQAKENHMQVLLGNTDFGSGSEYNQALGVKRTLSVKNALVKGVEKMDIKTISFGETK 156
 DB 62 VQAKENHMQVLLGNTDFGSGSEYNQALGVKRTLSVKNALVKGVEKMDIKTISFGETK 121
 QY 157 PKCAQKTRCYKRNRRVDVVKLMK 179
 DB 122 PKCTQKTRCYKRNRRVDVVKLVK 144
 RESULT 9
 AAW89983
 ID AAW89983 standard; Protein; 104 AA.
 AC AAW89983;
 XX 20-MAR-2003 (updated)
 DT 18-FEB-1999 (first entry)
 XX
 DE Expressed antigen for clone Y89A.
 XX
 KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.
 XX
 OS Helicobacter pylori.
 XX
 PN WO9849314-A2.
 XX
 PD 05-NOV-1998.
 XX
 PF 25-APR-1998; 98WO-US08487.
 XX
 PR 14-OCT-1997; 97US-0061958.
 PR 25-APR-1997; 97US-0045107.
 XX
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 XX Chow TP, Fry XE, Lim MY, McAtee CP;
 XX WPI; 1999-009433/01.
 XX
 FT New Helicobacter pylori antigens and related nucleic acid sequences
 FT - useful in serological diagnosis and protective vaccines, providing
 PT long-lasting immune response

XX Claim 16; Page 322; 402pp; English.

PS The present sequence represents a Helicobacter pylori antigenic protein

XX that is characterised by immunoreactivity with H. pylori-positive

CC antisera. The proteins are highly immunogenic and induce a long-lasting

CC immune response that persists even after antimicrobial treatment. In

CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are

CC highly sensitive and specific. The specification also describes 69

CC previously unrecognised immunogenic cluster families. H. pylori antigens

CC are used to detect H. pylori-specific antibodies, for diagnosing

CC infection or to confirm eradication of infection, and in vaccines to

CC protect against H. pylori infection and related diseases (gastritis,

CC peptic ulcer, gastric adenocarcinoma/lymphoma).

CC (Updated on 20-VAR-2003 to correct PF field.)

XX Sequence 104 AA;

XX Query Match 58.7%; Score 532; DB 20; Length 104;

XX Best Local Similarity 100.0%; Pred. No. 2.2e-44;

XX Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 IYDFDKYIKESQDETLDIIVQAKENHMQVLLGNTDFGSGSEYNQALGVKRTLSVKN 135

DB 1 IYDFDKYIKESQDETLDIIVQAKENHMQVLLGNTDFGSGSEYNQALGVKRTLSVKN 60

QY 136 ALVKGVEKDMIKTISFGETKPKCAQKTRCYKRNRRVDVVKMK 179

DB 61 ALVKGVEKDMIKTISFGETKPKCAQKTRCYKRNRRVDVVKMK 104

RESULT 10

AAV55090

ID AAV55090 standard; Protein; 172 AA.

XX AC AAV55090;

XX DT 01-MAR-2000 (first entry)

XX DE M. catarrhalis BASB019 protein sequence #2.

XX KW BASB019 protein; diagnosis; infectious organism; auditory nerve damage;

XX genetic mutation screening; antibody production; vaccine; otitis media;

XX bacterial infection; pneumonia; sinusitis; nosocomial infection;

XX invasive disease; delayed speech learning; bacteria adhesion prevention;

XX upper respiratory tract infection; middle ear infection; therapy.

XX OS Moraxella catarrhalis.

XX PN W09957277-A2.

XX PD 11-NOV-1999.

XX PF 03-MAY-1999; 99WO-EP03038.

XX PR 06-MAY-1998; 98GB-0009683.

XX PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Ruelle J;

XX WPI: 2000-062148/05.

XX DR N-PSDB: AAZ40352.

XX Novel BASB019 polynucleotides and polypeptides from Moraxella

PT catarrhalis used to prepare vaccines against bacterial infections

XX Claim 3; Fig 3; 101pp; English.

XX This sequence is a Moraxella catarrhalis BASB019 protein of the

CC invention. The sequences can be used for diagnosis of disease, staging of

CC disease, or determining response of an infectious organism to drugs. The

CC polynucleotides may be used as a source for hybridisation probes, and for

CC screening of genetic mutations, serotype, organism or strain

CC identification, identification of mutation in BASB013 sequences, and as

CC components of arrays which are useful for diagnostic and prognostic

CC purposes. The polypeptides can be used to produce antibodies, and as a

CC target for the screening of antimicrobial drugs. The polypeptides can

CC also be used in vaccine formulations, and to identify agonists and

CC antagonists. The polypeptides, antibodies, agonists and antagonists

CC (which are bacteriostatic) are used for the treatment and prevention of

CC diseases including bacterial infection, otitis media in infants and

CC children, pneumonia in the elderly, sinusitis, nosocomial infections and

CC invasive diseases, chronic otitis media with hearing loss, fluid

CC accumulation in the middle ear, auditory nerve damage, delayed speech

CC learning, infection of the upper respiratory tract and middle ear

CC infection. They are also used in the prevention of adhesion of bacteria

CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular

CC proteins on wounds, and to thus prevent tissue damage and/or block the

CC normal progression of pathogenesis in infections initiated other than by

CC the implantation of in-dwelling devices or by other surgical techniques.

CC The frequency of Moraxella catarrhalis infections has risen dramatically,

CC and it is no longer common to isolate M. catarrhalis strains that are

CC resistant to standard antibiotics. The BASB019 products of the invention

CC can be used screen for new antibacterial compounds that may target these

CC resistant bacteria.

XX Query Match 24.4%; Score 221.5; DB 21; Length 172;

XX Best Local Similarity 32.8%; Pred. No. 1.3e-13;

XX Matches 57; Conservative 34; Mismatches 62; Indels 21; Gaps 5;

QY 5 SVFSFLVAFLLVAGCGHKMDKNTVAGDVSATQVTPVTTPEAPEKEPAPVVEEK 64

DB 15 SVLTEM-----TGCANKSTQVMVAP-NAPTGYTVIYT-----GVAPLVND 56

QY 65 PAVE--SGTIIASIFDFDKYIKESQDETLDIIVQAKENH-MQVLLGNTDFGSGSEY 121

DB 57 ETVKALASTLPSLVYDFDSDIIRKQAAAILDEQAQFLTNTQATRVLVAGHTDGRGREY 116

QY 122 NQALGVKRTLSVKNALVIKVEKDMIKTISFGETKPKCAQKTRCYKRNRRVDV 175

DB 117 NWSLGERAVAVRNVLLGKNGINQASVEIISFGERPFAFGNEAWSONRAEL 170

RESULT 11

AAV55091

ID AAV55091 standard; Protein; 172 AA.

XX AC AAV55091;

XX DT 01-MAR-2000 (first entry)

XX DE M. catarrhalis BASB019 protein sequence #3.

XX KW BASB019 protein; diagnosis; infectious organism; auditory nerve damage;

XX genetic mutation screening; antibody production; vaccine; otitis media;

XX bacterial infection; pneumonia; sinusitis; nosocomial infection;

XX invasive disease; delayed speech learning; bacteria adhesion prevention;

XX upper respiratory tract infection; middle ear infection; therapy.

XX OS Moraxella catarrhalis.

XX PN W09957277-A2.

XX PD 11-NOV-1999.

XX PF 03-MAY-1999; 99WO-EP03038.

XX PR 06-MAY-1998; 98GB-0009683.

XX PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Ruelle J;

DR WPI; 2000-062148/05.
 DR N-PSDB; AA240353.
 XX
 XX Novel BASB019 polynucleotides and polypeptides from Moraxella
 PT catarrhalis used to prepare vaccines against bacterial infections
 XX
 XX Claim 3; Fig 3; 101pp; English.
 PS
 XX This sequence is a Moraxella catarrhalis BASB019 protein of the
 CC invention. The sequences can be used for diagnosis of disease, staging of
 CC disease, or determining response of an infectious organism to drugs. The
 CC polynucleotides may be used as a source for hybridisation probes, and for
 CC screening of genetic mutations, serotype, organism or strain
 CC identification, identification of mutation in BASB013 sequences, and as
 CC components of arrays which are useful for diagnostic and prognostic
 CC purposes. The polypeptides can be used to produce antibodies, and as a
 CC target for the screening of antimicrobial drugs. The polypeptides can
 CC also be used in vaccine formulations, and to identify agonists and
 CC antagonists. The polypeptides, antibodies, agonists and antagonists
 CC (which are bacteriostatic) are used for the treatment and prevention of
 CC diseases including bacterial infection, otitis media in infants and
 CC children, pneumonia in the elderly, sinusitis, nosocomial infections and
 CC invasive diseases, chronic otitis media with hearing loss, fluid
 CC accumulation in the middle ear, auditive nerve damage, delayed speech
 CC learning, infection of the upper respiratory tract and middle ear
 CC infection. They are also used in the prevention of adhesion of bacteria
 CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
 CC proteins on wounds, and to thus prevent tissue damage and/or block the
 CC normal progression of pathogenesis in infections initiated other than by
 CC the implantation of in-dwelling devices or by other surgical techniques.
 CC The frequency of Moraxella catarrhalis infections has risen dramatically,
 CC and it is no longer common to isolate M. catarrhalis strains that are
 CC resistant to standard antibiotics. The BASB019 products of the invention
 CC can be used screen for new antibacterial compounds that may target these
 CC resistant bacteria.
 XX
 SQ Sequence 172 AA;
 Query Match 24.3%; Score 220.5; DB 21; Length 172;
 Best Local Similarity 32.3%; Pred. No. 1.7e-13;
 Matches 56; Conservative 35; Mismatches 62; Indels 21; Gaps 5;
 QY 5 SVFSFLVAFLLVAGCSHKMDNKTVAGDVSAKTQTAPVTPTEPAPEKEEPKQBPAPVVEEK 64
 DB 15 SVLTFM-----TGCANKSTSQVMVAP--NAPGTGTVIYT-----GVAPLVND 56
 QY 65 PAVE--SGTIIASIVDFDKYIKESDQETLDELIVQAKENH-MQVLLEGNTDFGSGREY 121
 DB 57 ETIKTLASTLPSLVYDFDPSDEIKPQAAAILDQEQAFLLTNTQTARVLVAGHTDERSREY 116
 QY 122 NOALGVKRTLVSKNALVIGVEKMDIKTISFGETPKCAQKTRCYKERNRVDV 175
 DB 117 NWSLGERAVAVRNLLSGKINQASVEIISFGERPFAFGTNEAWSQNRRAEL 170
 RESULT 12
 ID AAY55092
 XX AAY55092 standard; Protein; 172 AA.
 AC AAY55092;
 XX
 XX 01-MAR-2000 (first entry)
 XX
 XX M. catarrhalis BASB019 protein sequence #4.
 XX
 KW BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
 KW Genetic mutation screening; antibody production; vaccine; otitis media;
 KW bacterial infection; pneumonia; sinusitis; nosocomial infection;
 KW invasive disease; delayed speech learning; bacteria adhesion prevention;
 KW upper respiratory tract infection; middle ear infection; therapy.
 XX
 OS Moraxella catarrhalis.
 XX

PN WO9957277-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 03-MAY-1999; 99WO-EP03038.
 XX
 PR 06-MAY-1998; 98GB-0009683.
 XX
 XX (SMIK) SMITHLINE BEECHAM BIOLOGICALS.
 PA
 XX Ruelle J;
 PI
 XX WPI; 2000-062148/05.
 DR N-PSDB; AA240354.
 DR
 PT Novel BASB019 polynucleotides and polypeptides from Moraxella
 PT catarrhalis used to prepare vaccines against bacterial infections
 XX
 PS Claim 3; Fig 3; 101pp; English.
 XX
 CC This sequence is a Moraxella catarrhalis BASB019 protein of the
 CC invention. The sequences can be used for diagnosis of disease, staging of
 CC disease, or determining response of an infectious organism to drugs. The
 CC polynucleotides may be used as a source for hybridisation probes, and for
 CC screening of genetic mutations, serotype, organism or strain
 CC identification, identification of mutation in BASB013 sequences, and as
 CC components of arrays which are useful for diagnostic and prognostic
 CC purposes. The polypeptides can be used to produce antibodies, and as a
 CC target for the screening of antimicrobial drugs. The polypeptides can
 CC also be used in vaccine formulations, and to identify agonists and
 CC antagonists. The polypeptides, antibodies, agonists and antagonists
 CC (which are bacteriostatic) are used for the treatment and prevention of
 CC diseases including bacterial infection, otitis media in infants and
 CC children, pneumonia in the elderly, sinusitis, nosocomial infections and
 CC invasive diseases, chronic otitis media with hearing loss, fluid
 CC accumulation in the middle ear, auditive nerve damage, delayed speech
 CC learning, infection of the upper respiratory tract and middle ear
 CC infection. They are also used in the prevention of adhesion of bacteria
 CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
 CC proteins on wounds, and to thus prevent tissue damage and/or block the
 CC normal progression of pathogenesis in infections initiated other than by
 CC the implantation of in-dwelling devices or by other surgical techniques.
 CC The frequency of Moraxella catarrhalis infections has risen dramatically,
 CC and it is no longer common to isolate M. catarrhalis strains that are
 CC resistant to standard antibiotics. The BASB019 products of the invention
 CC can be used screen for new antibacterial compounds that may target these
 CC resistant bacteria.
 XX
 SQ Sequence 172 AA;
 Query Match 24.1%; Score 218.5; DB 21; Length 172;
 Best Local Similarity 33.3%; Pred. No. 2.6e-13;
 Matches 58; Conservative 29; Mismatches 66; Indels 21; Gaps 5;
 QY 5 SVFSFLVAFLLVAGCSHKMDNKTVAGDVSAKTQTAPVTPTEPAPEKEEPKQBPAPVVEEK 64
 DB 15 SVLTFM-----TGCANKSTSQVMVAP--NAPGTGAGVITGVAPLVND 56
 QY 65 PAVE--SGTIIASIVDFDKYIKESDQETLDELIVQAKENH-MQVLLEGNTDFGSGREY 121
 DB 57 ETIKTLASTLPSLVYDFDPSDEIKPQAAAILDQEQAFLLTNTQTARVLVAGHTDERSREY 116
 QY 122 NOALGVKRTLVSKNALVIGVEKMDIKTISFGETPKCAQKTRCYKERNRVDV 175
 DB 117 NWSLGERAVAVRNLLSGKINQASVEIISFGERPFAFGTNEAWSQNRRAEL 170
 RESULT 13
 ID AAY55089
 XX AAY55089 standard; Protein; 172 AA.
 AC AAY55089;
 XX

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:31:22 ; Search time 11.2461 Seconds
(without alignments)
673.448 Million cell updates/sec

Title: US-10-080-113-2
Perfect score: 906
Sequence: 1 MKRSSVFLVAFLLVAGCS.....AOKTRECYNERNVDVVKLMK 179

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata1/iaa/5A COMB.pdp.*
2: /cgn2_6/prodata1/iaa/5B COMB.pdp.*
3: /cgn2_6/prodata1/iaa/6A COMB.pdp.*
4: /cgn2_6/prodata1/iaa/6B COMB.pdp.*
5: /cgn2_6/prodata1/iaa/6C COMB.pdp.*
6: /cgn2_6/prodata1/iaa/backfiles1.pdp.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	204	22.5	136	US-09-328-352-6310	Sequence 6310, Ap
2	191	21.1	235	US-09-328-352-6310	Sequence 19595, A
3	189	20.9	187	US-09-328-352-6310	Sequence 29582, A
4	183	20.2	153	5173294-2	Patent No. 5173294
5	169	18.7	335	US-08-836-500A-2	Sequence 2, Appl
6	161.5	17.8	266	US-09-328-352-6310	Sequence 18046, A
7	160	17.7	192	US-09-198-452A-830	Sequence 830, App
8	153	16.9	161	US-08-572-447C-11	Sequence 11, Appl
9	153	16.9	161	US-09-267-747-11	Sequence 11, Appl
10	153	16.9	226	US-08-572-447C-15	Sequence 15, Appl
11	153	16.9	226	US-09-267-747-15	Sequence 15, Appl
12	153	16.9	351	US-09-328-352-6310	Sequence 30094, A
13	151	16.7	214	US-08-572-447C-13	Sequence 13, Appl
14	151	16.7	214	US-09-267-747-13	Sequence 13, Appl
15	149.5	16.5	259	US-09-328-352-6310	Sequence 5775, Ap
16	144	15.9	379	US-09-328-352-5219	Sequence 5219, Ap
17	135.5	15.0	359	US-08-457-997B-2	Sequence 2, Appl
18	135.5	15.0	359	US-08-467-722A-2	Sequence 2, Appl
19	135.5	15.0	359	US-09-451-184-2	Sequence 2, Appl
20	130	14.3	257	US-09-328-352-6621	Sequence 6621, Ap
21	128	14.1	417	US-09-328-352-6621	Sequence 26245, A
22	127.5	14.1	358	US-08-310-394-1	Sequence 1, Appl
23	126	13.9	278	US-09-328-352-5824	Sequence 5824, Ap
24	114	12.6	272	US-09-328-352-5824	Sequence 29681, A
25	113	12.5	169	US-09-328-352-6870	Sequence 19373, A
26	104.5	11.5	472	US-09-328-352-6870	Sequence 6870, Ap
27	101.5	11.2	753	US-09-328-352-6870	Sequence 17612, A

Sequence 27792, A
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 32, Appl
Sequence 5207, Ap
Sequence 72, Appl
Sequence 19600, A
Sequence 3, Appl
Sequence 4, Appl
Sequence 12, Appl
Patent No. 5284931
Sequence 11, Appl
Sequence 1, Appl
Sequence 15, Appl
Sequence 19, Appl
Sequence 1, Appl

US-09-252-991A-27792
US-08-072-610-2
US-08-719-822B-2
US-09-092-458-2
US-08-714-741-32
US-09-107-532A-5207
US-08-235-836C-72
US-09-252-991A-19600
US-09-240-915-3
US-09-591-435-3
US-08-515-251A-4
US-08-989-299-12
US-08-931-3
US-08-425-389B-11
US-08-318-039A-1
US-08-318-038D-15
US-08-237-496C-19
US-08-435-568A-1

ALIGNMENTS

RESULT 1
US-09-328-352-6310
; Sequence 6310, Application US/09328352
; Patent No. 6562558
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6310
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6310

Query Match 22.5%; Score 204; DB 4; Length 196;
Best Local Similarity 30.4%; Pred. No. 2.6e-13;
Matches 52; Conservative 32; Mismatches 83; Indels 4; Gaps 3;

QY 7 FSLVAFLLVAGC-SHKMDNKTVAGDVS AKTVQAPVTTEPAPEKEPKQBPAPWEEKP 65
Db 18 FLLSAAVMTGCSARKEPATTTATTGTPSTVNTGTSDDLALNAQNLGASSKGVTE-- 75
QY 66 AVESGTTIATSYFDFDKYEIKESQDTLDEIVQAKEN-HMQVLLGNNTDFGSEYNQA 124
Db 76 ANKAALAKRVVHFDYDSDSLSTEDYQLQAHAQFLMANANSKVALTGHTDERTREYNMA 135
QY 125 LGVKTSLVKALVIKGVKMDIKTISFGETPKPKCAQKTRCYKENRVDV 175
Db 136 LGERRAKAVNYLITSGVNPQCLEAVSYGKEAPVNPQGHDESAWKENRVEI 186

RESULT 2
US-09-252-991A-19595
; Sequence 19595, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107136.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27

```
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19595
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19595

Query Match
Best Local Similarity 21.1%; Score 191; DB 4; Length 235;
Matches 48; Conservative 27; Mismatches 58; Indels 12; Gaps 4;

QY 38 QTAPVTTEPAPEKEPKQEPAPVVEKPAVESGTIIASIVFDFDKYIKESDOETLDEIV 97
DB 99 QVAP-----PPQPVEEAVPPPPVKEETIV-----VRDLHFAFDSKKVDAADSEKLGIA 149

QY 98 QKAKEN--HMQVLEGNITDEFGSSEYNQALGVKRTLSVKNALVIKGVKDMIKTI-SFGE 154
DB 150 ERLKGEAASTRLSITGHTDSVGSAYNQKLSERRANAVANYLIDAGVPSSIVGVQGLGE 209

QY 155 TKPKCAQKTRCYKENRRVDVKLMK 179
DB 210 SQPVADNKTREGRAENRRVEILIKR 234

RESULT 3
US-09-252-991A-29582
; Sequence 29582, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29582
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29582

Query Match
Best Local Similarity 20.9%; Score 189; DB 4; Length 187;
Matches 51; Conservative 31; Mismatches 75; Indels 26; Gaps 4;

QY 1 MKRSVSFSLVAFLLVAGCSHKMDNKTIVAGDVSAKTIVQTAPVTTEPAPEKEPKQEPAPV 60
DB 23 LKFGKFAALAMAVAVGCSK-----GGDASG-----EGANGGVDFN---AGY 63

QY 61 VEKPAVESG-----TIIASIVFDFDKYIKESDOETLDEIVQKAKENHMOVLLEGN 113
DB 64 GANSGVDGSLSDAALRAITTFEYDSSDLKPEAKRALDVHAKDLKSGQGVLEGT 123

QY 114 DEFGSSEYNQALGVKRTLSVKNALVIKGVKDMIKTISFGETPKCAQKTRCYKENRRV 173
DB 124 DERGTREYNALGERRAKQVRLVQGVSPAQLVLSYGVKRPVATGHDHESWAHNRV 183

QY 174 DVK 176
DB 184 ELK 186

RESULT 4
5173294-2
; Patent No. 5173294
; APPLICANT: MURPHY, TIMOTHY F.; APICELLA, MICHAEL A.
; TITLE OF INVENTION: DNA PROBE FOR THE IDENTIFICATION
; OF HAEMOPHILUS INFLUENZAE
```

```
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/92,948
; FILING DATE: 08-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 932,872
; FILING DATE: 18-NOV-1986
; SEQ ID NO: 2
; LENGTH: 153
5173294-2

Query Match
Best Local Similarity 20.2%; Score 183; DB 6; Length 153;
Matches 58; Conservative 17; Mismatches 49; Indels 58; Gaps 6;

QY 11 VAFLLVAG-----CSHKMDNKTIVAGDVSAKTIVQTAPVTTEPAPEKEPKQEPAPVVEE 63
DB 5 VKSLVAGSVAALACS--SSNDAAGNAAQTF----- 36

QY 64 KPAVESGTIIA-----SIVFDFDKYIKESDOETLD-----EIVQKAKENHMOVLE 110
DB 37 -----GGYSVADLQORYNTVTFGDKYDITGEYQILDAAHYALNATPAK-----VIVE 86

QY 111 GNTDEFGSSEYNQALGVKRTLSVKNALVIKGVKDMIKTISFGETPKCAQKTRCYKEN 170
DB 87 GNTDERGTPEYNIALQORRADAVKYLAKGVDAKGLGTVSYGEEKPAVLGHDEAAYSKN 146

QY 171 RR 172
DB 147 RR 148

RESULT 5
US-08-836-500A-2
; Sequence 2, Application US/08836500A
; Patent No. 6197929
; GENERAL INFORMATION:
; APPLICANT: Binz, Hans
; APPLICANT: Bausant, Thierry
; APPLICANT: Haeuw, Jean-Francois
; APPLICANT: Nguyen Ngoc, Thien
; TITLE OF INVENTION: Carrier Protein Having an Adjuvant
; TITLE OF INVENTION: Effect, Immunogenic Complex Containing It, Process for
; TITLE OF INVENTION: Their Preparation, Nucleotide Sequence and Vaccines
; Patent No. 6197929
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rocket, Milnamow & Katz, Ltd.
; STREET: 180 N. Stetson, 2 Prudential Plaza, Suite
; STREET: 4700
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,500A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Katz, Martin L.
; REGISTRATION NUMBER: 25,011
; REFERENCE/DOCKET NUMBER: PIE1514F0180US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-616-5400
; TELEFAX: 312-616-5460
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 335 amino acids
```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-836-500A-2

Query Match 18.7%; Score 169; DB 3; Length 335;
Best Local Similarity 27.6%; Pred. No. 2.3e-09;
Matches 54; Conservative 31; Mismatches 87; Indels 24; Gaps 4;
QY 4 SSVFSLVAFLLVAGGSHKMDNKTVDGVS AKTVQAP-----VTEPAPEKEEPKQ 55
DB 130 SPVAGVEWAVTRDIATRELYQWVNNIGDAGTGTDPNGMLSLGVS YRFQEDAAPV 189
QY 56 EPAPVVEEKPAVESGTIIASIFDPDKYIKESDQETLDEIVQK-----AKENHMQVLLE 110
DB 190 APAPAPAEVATKHTFLKSDVLFENKATLKPEGQALDQLYTQLSNMPPKDGSAVWL-- 247
QY 111 GNTDEGSSEYNCALGVKRTLSVKNALVKGVEKDMIKTISGETXP-----KCAQ 161
DB 248 GYTRIGSEAYNQOLSEKRAQSVVDYLVAKGIPAGKISARGMGESNPVTGNTCDNVKARA 307
QY 162 KTRCYKENRRVDVKL 177
DB 308 ALIDCLAPDRRVEIV 323

RESULT 6

US-09-252-991A-18046
; Sequence 18046, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18046
; LENGTH: 266
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18046

Query Match 17.8%; Score 161.5; DB 4; Length 266;
Best Local Similarity 31.5%; Pred. No. 9.6e-09;
Matches 35; Conservative 26; Mismatches 47; Indels 3; Gaps 2;
QY 70 GTII--ASYFPDPKYIKESDQETLDEIVQKAKEN-HMQVLLEGNTDFGSSYVQALG 126
DB 141 GTVTFGDFLFDLKDGLKFGAMRN:QQLAEFLQQNPERQVIGVGYTDTGTSANYNQRLS 200
QY 127 VKRTLSVKNALVKGVEKDMIKTISFGETKPKCAQKTRCYKENRRVDVKL 177
DB 201 ERRADSVRMALLSGISPERVATRGYGYKPYVASNGTSSGRAMRRVEVTI 251

RESULT 7

US-09-198-452A-830
; Sequence 830, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24

; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 830
; LENGTH: 192
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-830

Query Match 17.7%; Score 160; DB 4; Length 192;
Best Local Similarity 33.3%; Pred. No. 8.7e-09;
Matches 40; Conservative 26; Mismatches 50; Indels 4; Gaps 3;
QY 62 BEKPAVSGTIIA--SIYFDPKYIK-ESDQETLDEIVQKAKENHMQVL-LEGNTEBFG 117
DB 70 KEKQYKSSQVAAPRNITFATDSYTIKGEENLAILTNLVHYMKKNPKATLVIEGHTDBRG 129
QY 118 SSEYNQALGVKRTLSVKNALVKGVEKDMIKTISFGETKPKCAQKTRCYKENRRVDVKL 177
DB 130 AASYNLALGARRAIAKEHLRKQGISADRLSTISYKHEHPLNSOHNELAWQONRRTPEKI 189

RESULT 8

US-08-572-447C-11
; Sequence 11, Application US/08572447C
; Patent No. 5955090
; GENERAL INFORMATION:
; APPLICANT: Knapp, Bernhard
; APPLICANT: Hungerer, Klaus-Dieter
; APPLICANT: Broker, Michael
; APPLICANT: Von Specht, Bernd-Ulrich
; APPLICANT: Domdey, Horst
; TITLE OF INVENTION: Immunogenic Hybrid Protein OprF-OprI
; TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/572,447C
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94120023.0
; FILING DATE: 16-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: M. Paul Barker
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 05552.1395-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-572-447C-11

Query Match 16.9%; Score 153; DB 2; Length 161;
Best Local Similarity 28.0%; Pred. No. 3.6e-08;
Matches 46; Conservative 26; Mismatches 76; Indels 16; Gaps 4;
QY 16 VAGSHKMDNKTVDGVS AKTVQAPVTEPAPEKEEPKQEPAPVVEKPAVESGTIIAS 75

Db 7 VADVCSDDNDGVCNDVDCPDTPANVTVD-----ANGCPAAVEVRVQ-----LD 52
Qy 76 IYFDFKYEIKESQOETLDEIVQKAKE-NHMQVLLEGTDFGSGSEYNQALGVKRTLSVK 134
Db 53 VKFDFKSKVKENSADIKNLADFMKQYPSSTTVEGHTDSVGTDAYNQKLSERRANVR 112
Qy 135 NALVIK-GVEKDMIKTISFGTKPKCAQKRECYKENRRVDVKL 177
Db 113 DVLVNEYGVEGRVNAVGYGESRPVADNATAEGRAINRVEAEV 156
RESULT 9
US-09-267-747-11
; Sequence 11, Application US/09267747
; Patent No. 6300102
; GENERAL INFORMATION:
; APPLICANT: Knapp, Bernhard
; APPLICANT: Hungerer, Klaus-Dieter
; APPLICANT: Broker, Michael
; APPLICANT: Von Specht, Bernd-Ulrich
; APPLICANT: Domdey, Horst
; TITLE OF INVENTION: Immunogenic Hybrid Protein OprF-OprI
; TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/267,747
; FILING DATE: 14-DEC-1995
; CLASSIFICATION:
; APPLICATION NUMBER: US/08/572,447
; FILING DATE: 14-DEC-1995
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-267-747-11
Query Match 16.9%; Score 153; DB 4; Length 161;
Best Local Similarity 28.0%; Pred. No. 3.6e-08;
Matches 46; Conservative 26; Mismatches 76; Indels 16; Gaps 4;
Qy 16 VAGCSHKMDNKTVAGDYSAKTVQTPVTEPAKEBPKEPAPVVEKPAVESGIIAS 75
Db 7 VADVCSDDNDGVCNDVDCPDTPANVTVD-----ANGCPAAVEVRVQ-----LD 52
Qy 76 IYFDFKYEIKESQOETLDEIVQKAKE-NHMQVLLEGTDFGSGSEYNQALGVKRTLSVK 134
Db 53 VKFDFKSKVKENSADIKNLADFMKQYPSSTTVEGHTDSVGTDAYNQKLSERRANVR 112

Qy 135 NALVIK-GVEKDMIKTISFGTKPKCAQKRECYKENRRVDVKL 177
Db 113 DVLVNEYGVEGRVNAVGYGESRPVADNATAEGRAINRVEAEV 156
RESULT 10
US-08-572-447C-15
; Sequence 15, Application US/08572447C
; Patent No. 5955090
; GENERAL INFORMATION:
; APPLICANT: Knapp, Bernhard
; APPLICANT: Hungerer, Klaus-Dieter
; APPLICANT: Broker, Michael
; APPLICANT: Von Specht, Bernd-Ulrich
; APPLICANT: Domdey, Horst
; TITLE OF INVENTION: Immunogenic Hybrid Protein OprF-OprI
; TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/572,447C
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; APPLICATION NUMBER: EP 94120023.0
; FILING DATE: 16-DEC-1994
; NAME: M. Paul Barker
; ATTORNEY/AGENT INFORMATION:
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 05552.1395-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 226 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-572-447C-15
Query Match 16.9%; Score 153; DB 2; Length 226;
Best Local Similarity 28.0%; Pred. No. 5.7e-08;
Matches 46; Conservative 26; Mismatches 76; Indels 16; Gaps 4;
Qy 16 VAGCSHKMDNKTVAGDYSAKTVQTPVTEPAKEBPKEPAPVVEKPAVESGIIAS 75
Db 72 VADVCSDDNDGVCNDVDCPDTPANVTVD-----ANGCPAAVEVRVQ-----LD 117
Qy 76 IYFDFKYEIKESQOETLDEIVQKAKE-NHMQVLLEGTDFGSGSEYNQALGVKRTLSVK 134
Db 118 VKFDFKSKVKENSADIKNLADFMKQYPSSTTVEGHTDSVGTDAYNQKLSERRANVR 177
Qy 135 NALVIK-GVEKDMIKTISFGTKPKCAQKRECYKENRRVDVKL 177
Db 178 DVLVNEYGVEGRVNAVGYGESRPVADNATAEGRAINRVEAEV 221
RESULT 11
US-09-267-747-15

; Sequence 15, Application US/09267747
; Patent No. 6300102
; GENERAL INFORMATION:
; APPLICANT: Knapp, Bernhard
; APPLICANT: Hungerer, Klaus-Dieter
; APPLICANT: Broker, Michael
; APPLICANT: Von Specht, Bernd-Ulrich
; APPLICANT: Domdey, Horst
; TITLE OF INVENTION: Immunogenic Hybrid Protein OprF-OprI
; TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/267,747
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/572,447
; FILING DATE: 14-DEC-1995
; APPLICATION NUMBER: EP 94120023.0
; FILING DATE: 16-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: M. Paul Barker
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 05552.1395-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 226 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-267-747-15

Query Match 16.9%; Score 153; DB 4; Length 226;
Best Local Similarity 28.0%; Pred. No. 5.7e-08;
Matches 46; Conservative 26; Mismatches 76; Indels 16; Gaps 4;

QY 16 VAGCSHKMDNKTAVGDPVSAKTQVPTTTPAPEKEPKQEPAPVVEKPAVESGTTIAS 75
DB 72 VADVCSDDNDGVCNDVKCPDTPANTVD-----ANGCPAAVEVRVQ-----LD 117

QY 76 IYDFDKYEKESQDTLDEIVQAKB-NHMVLLGNTDFGSSYNQALGVKRTLSVK 134
DB 118 VKFDFDKSKVKNESYADIKNLADFMKQYPSSTTTVEGHTDSVGTDAYNQLSERRANAVR 177

QY 135 NALVIK-GVEKDMIKTISFGETKPKCAQKTRCYKENRRVDVKL 177
DB 178 DVLVNEYGVGGVRNAVGYGESRPVADNATAEGRAINRRVEAEV 221

RESULT 12
US-09-252-991A-30094
; Sequence 30094, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30094
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30094

Query Match 16.9%; Score 153; DB 4; Length 351;
Best Local Similarity 28.0%; Pred. No. 1.1e-07;
Matches 46; Conservative 26; Mismatches 76; Indels 16; Gaps 4;

QY 16 VAGCSHKMDNKTAVGDPVSAKTQVPTTTPAPEKEPKQEPAPVVEKPAVESGTTIAS 75
DB 197 VADVCSDDNDGVCNDVKCPDTPANTVD-----ANGCPAAVEVRVQ-----LD 242

QY 76 IYDFDKYEKESQDTLDEIVQAKB-NHMVLLGNTDFGSSYNQALGVKRTLSVK 134
DB 243 VKFDFDKSKVKNESYADIKNLADFMKQYPSSTTTVEGHTDSVGTDAYNQLSERRANAVR 302

QY 135 NALVIK-GVEKDMIKTISFGETKPKCAQKTRCYKENRRVDVKL 177
DB 303 DVLVNEYGVGGVRNAVGYGESRPVADNATAEGRAINRRVEAEV 346

RESULT 13
US-08-572-447C-13
; Sequence 13, Application US/08572447C
; Patent No. 5955090
; GENERAL INFORMATION:
; APPLICANT: Knapp, Bernhard
; APPLICANT: Hungerer, Klaus-Dieter
; APPLICANT: Broker, Michael
; APPLICANT: Von Specht, Bernd-Ulrich
; APPLICANT: Domdey, Horst
; TITLE OF INVENTION: Immunogenic Hybrid Protein OprF-OprI
; TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/572,447C
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94120023.0
; FILING DATE: 16-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: M. Paul Barker
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 05552.1395-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:

LENGTH: 214 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-572-447C-13

Query Match 16.7%; Score 151; DB 2; Length 214;
Best Local Similarity 28.6%; Pred. No. 8.6e-08;
Matches 46; Conservative 24; Mismatches 75; Indels 16; Gaps 4;
QY 16 VAGCSHKMDNKTAVGDSVAKTVQAPVTTPEPAKEEPKQEPAPVVEEKPAVESGTTIAS 75
DB 5 VADVCSDSNDGVCNDVDCPDTPANVTVD-----ANGCPAAVEVVRVQ-----LD 50
QY 76 IYFDFDKVEIKESDQETLDEIVQAKE-NEMQVLLGNTDFGSSSEYNOALGVKRTLSVK 134
DB 51 VKFDFDKSVKENSYADIKNLADFMKQYPSTSTTVEGHTDSVGTDAYNKLSERRANAVR 110
QY 135 NALVIK-GVEKDMIKTISFGETKPKCAQKTRCYKERNRVD 174
DB 111 DVLVNEYGVEGRVNAVGYGESRPVADNATAEGRAINRVE 151

RESULT 14
US-09-267-747-13
Sequence 13, Application US/09267747
Patent No. 6300102
GENERAL INFORMATION:
APPLICANT: Knapp, Bernhard
APPLICANT: Hungerer, Klaus-Dieter
APPLICANT: Broker, Michael
APPLICANT: Von Specht, Bernd-Ulrich
APPLICANT: Domdey, Horst
TITLE OF INVENTION: Immunogenic Hybrid Protein Oprf-OprI
DERIVED FROM Pseudomonas aeruginosa Membrane Proteins.
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: /09/267,747
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/572,447
FILING DATE: 14-DEC-1995
APPLICATION NUMBER: EP 94120023.0
FILING DATE: 16-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: M. Paul Barker
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 05552.1395-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 214 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-267-747-13

Query Match 16.7%; Score 151; DB 4; Length 214;
Best Local Similarity 28.6%; Pred. No. 9.6e-08;
Matches 46; Conservative 24; Mismatches 75; Indels 16; Gaps 4;
QY 16 VAGCSHKMDNKTAVGDSVAKTVQAPVTTPEPAKEEPKQEPAPVVEEKPAVESGTTIAS 75
DB 5 VADVCSDSNDGVCNDVDCPDTPANVTVD-----ANGCPAAVEVVRVQ-----LD 50
QY 76 IYFDFDKVEIKESDQETLDEIVQAKE-NEMQVLLGNTDFGSSSEYNOALGVKRTLSVK 134
DB 51 VKFDFDKSVKENSYADIKNLADFMKQYPSTSTTVEGHTDSVGTDAYNKLSERRANAVR 110
QY 135 NALVIK-GVEKDMIKTISFGETKPKCAQKTRCYKERNRVD 174
DB 111 DVLVNEYGVEGRVNAVGYGESRPVADNATAEGRAINRVE 151

RESULT 15
US-09-328-352-5775
Sequence 5775, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 5775
LENGTH: 259
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-5775

Query Match 16.5%; Score 149.5; DB 4; Length 259;
Best Local Similarity 35.6%; Pred. No. 1.6e-07;
Matches 37; Conservative 16; Mismatches 50; Indels 1; Gaps 1;
QY 75 SIYFDFDKYEIKESDQETLDEIVQK-AKENHMQVLLGNTDFGSSSEYNOALGVKRTLSV 133
DB 152 SITFDTNKSNIKPNFYATLDKVAQTLAEDNKSALVTGYDTNTGDSINIPLSQARQSV 211
QY 134 KNALVIKGVKDMIKTISFGETKPKCAQKTRCYKERNRVDVKL 177
DB 212 KNYLAGKGVPSRIDAQGYSSNFIADNSTASGREGQNRVEISI 255

Search completed: December 10, 2003, 18:38:37
Job time : 12.2461 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:33:52 ; Search time 22.0236 Seconds
(without alignments)
1511.612 Million cell updates/sec

Title: US-10-080-113-2

Perfect score: 906

Sequence: 1 MKRSVSFSLVAFLLVAGCS.....AQKTRCYKENVKRVVDVVKLMK 179

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	906	100.0	179	US-10-080-113-2	Sequence 2, Appli
2	169	18.7	344	US-10-169-953-2	Sequence 2, Appli
3	162	17.9	341	US-10-336-840-5	Sequence 5, Appli
4	162	17.9	341	US-10-336-840-8	Sequence 8, Appli
5	162	17.9	341	US-10-336-840-11	Sequence 11, Appli
6	162	17.9	344	US-10-336-840-1	Sequence 1, Appli
7	162	17.9	344	US-10-336-840-2	Sequence 2, Appli
8	162	17.9	344	US-10-336-840-4	Sequence 4, Appli
9	162	17.9	344	US-10-336-840-10	Sequence 10, Appli
10	160	17.7	192	US-09-747-348-2	Sequence 2, Appli
11	160	17.7	192	US-10-334-137-2	Sequence 2, Appli
12	155	17.1	344	US-10-336-840-3	Sequence 3, Appli
13	155	17.1	344	US-10-336-840-6	Sequence 6, Appli
14	155	17.1	344	US-10-336-840-7	Sequence 7, Appli
15	155	17.1	344	US-10-336-840-9	Sequence 9, Appli

Query Match 100.0%; Score 906; DB 12; Length 179;
Best Local Similarity 100.0%; Pred. No. 4.1e-90;
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRSVSFSLVAFLLVAGCSHKMDKTVAGDVSAKTQTAPVTTPAPEKEEPQPEAPV 60
DB 1 MKRSVSFSLVAFLLVAGCSHKMDKTVAGDVSAKTQTAPVTTPAPEKEEPQPEAPV 60
QY 61 VEEKPAVESGTTIIASIVFPDKYIKESDQBTLEIVQKAKENHMOVLLENTDFGSGS 120
DB 61 VEEKPAVESGTTIIASIVFPDKYIKESDQBTLEIVQKAKENHMOVLLENTDFGSGS 120
QY 121 YNQALGVKRTLSVKNALVIKGVKMDLKTISFGTKPKCAQKTRCYKENVKRVVDVVKLMK 179
DB 121 YNQALGVKRTLSVKNALVIKGVKMDLKTISFGTKPKCAQKTRCYKENVKRVVDVVKLMK 179

RESULT 2

ALIGNMENTS

RESULT 1
US-10-080-113-2
; Sequence 2, Application US/10080113
; Publication No. US20030166027A1
; GENERAL INFORMATION:
; APPLICANT: SACHS, GEORGE
; APPLICANT: VOLAND, PETRA
; TITLE OF INVENTION: COMPOSITIONS, TEST KITS AND METHODS FOR DETECTING
; FILE REFERENCE: 626 06 PA
; CURRENT APPLICATION NUMBER: US/10/080,113
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-10-080-113-2

Sequence 12, Appli
Sequence 37, Appli
Sequence 9, Appli
Sequence 38, Appli
Sequence 784, App
Sequence 492, App
Sequence 449, App
Sequence 3, Appli
Sequence 10782, A
Sequence 375, App
Sequence 26, Appli
Sequence 4, Appli
Sequence 24, Appli
Sequence 128, App
Sequence 228, App
Sequence 81, Appli
Sequence 932, App
Sequence 15, Appli
Sequence 2, Appli
Sequence 6891, Ap
Sequence 18, Appli
Sequence 11, App
Sequence 2, Appli
Sequence 63, Appli
Sequence 141, App
Sequence 8, Appli
Sequence 48, Appli
Sequence 6147, Ap
Sequence 7539, Ap

16 153 16.9 344 12 US-10-336-840-12
17 145.5 16.1 352 12 US-10-336-840-37
18 145.5 16.1 353 15 US-10-203-942-9
19 99.5 11.0 356 12 US-10-336-840-38
20 97 10.7 550 12 US-10-238-075-784
21 96.5 10.7 862 12 US-10-339-783-492
22 96.5 10.7 910 15 US-10-177-293-449
23 96 10.6 390 12 US-10-283-024-3
24 92 10.2 226 15 US-10-156-761-10782
25 90 9.9 407 9 US-09-764-870-375
26 90 9.9 407 15 US-10-125-540-375
27 89.5 9.9 380 11 US-09-998-279-26
28 89.5 9.9 380 12 US-10-283-024-4
29 89.5 9.9 385 11 US-09-998-279-24
30 89 9.8 257 12 US-09-882-227-128
31 89 9.8 285 10 US-09-881-752A-228
32 85.5 9.4 505 14 US-10-047-542-81
33 83.5 9.2 635 15 US-10-101-464A-932
34 83 9.2 680 7 US-08-469-583A-15
35 83 9.2 26926 10 US-09-759-508B-2
36 82.5 9.1 362 10 US-09-738-626-6891
37 82 9.1 507 7 US-08-469-583A-18
38 82 9.1 532 12 US-10-021-660-111
39 82 9.1 532 14 US-10-047-542-2
40 82 9.1 532 14 US-10-047-542-63
41 82 9.1 532 15 US-10-207-655-141
42 82 9.1 799 14 US-10-047-542-8
43 82 9.1 822 14 US-10-047-542-48
44 81 8.9 186 15 US-10-062-698-6147
45 81 8.9 714 12 US-10-032-585-7539

US-10-169-953-2
 ; Sequence 2, Application US/10169953
 ; Publication No. US20030044915A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Thierry BAUSSANT
 ; APPLICANT: Pascale JEANNIN
 ; APPLICANT: Yves DELNESTE
 ; APPLICANT: Françoise LAUNY
 ; APPLICANT: Jean-Yves BONNEFOY
 ; TITLE OF INVENTION: METHOD FOR PREPARING A POLYPEPTIDE SOLUBLE IN AN AQUEOUS SOLVENT
 ; FILE REFERENCE: D18390
 ; CURRENT APPLICATION NUMBER: US/10/169,953
 ; CURRENT FILING DATE: 2002-07-03
 ; PRIOR APPLICATION NUMBER: FR 00 00070
 ; PRIOR FILING DATE: 2000-01-04
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 344
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae
 ; FEATURE:
 ; OTHER INFORMATION: P40
 US-10-169-953-2

Query Match 18.7%; Score 169; DB 15; Length 344;
 Best Local Similarity 27.6%; Pred. No. 4e-08;
 Matches 54; Conservative 31; Mismatches 87; Indels 24; Gaps 4;
 QY 4 SSVFSLVAFLLVAGCSHKMDKTVAGDVSAGTVQTAP-----VTEPAPAKEEKPQ 55
 DB 139 SPVPAGGVEMAVTRDIATRLVYQVNNIGDAGTVGTDPNGMLSLGVSYRFGQDAAPVV 198
 QY 56 EPAPVVEKPAVESGTTIASIYFDPDKYIKESQDQETLDEIVK-----AKENHMOVLLE 110
 DB 199 APAPAPAEVATGFTLSDVLFNFNKRATLKPEGQQALDQLYTLQSNMDFKDGSAVVL-- 256
 QY 111 GNTDFGSSEYNOALGVKRTLSVKNALVIKGVKMDIKTISFGETKP-----KCAQ 161
 DB 257 GYTRIGSEAYNQSLSEKRAQSVVDYLVAKGIPAGKISARGMGSENPVTGNTCDNVKARA 316
 QY 162 KTRCYKENRRVDVKL 177
 DB 317 ALIDCLAPDRRVEIEV 332

RESULT 3
 US-10-336-840-5
 ; Sequence 5, Application US/10336840
 ; Publication No. US20030219454A1
 ; GENERAL INFORMATION:
 ; APPLICANT: TERRY, TAMSIN DEBORAH
 ; APPLICANT: TSENG, HSING-JU
 ; APPLICANT: HOBBS, RHONDA IVY
 ; APPLICANT: JENNINGS, MICHAEL PAUL
 ; APPLICANT: DOWNES, JOHN
 ; TITLE OF INVENTION: HAEMAGGLUTININ ANTIGEN
 ; FILE REFERENCE: 37955-0007
 ; CURRENT APPLICATION NUMBER: US/10/336,840
 ; CURRENT FILING DATE: 2003-01-06
 ; PRIOR APPLICATION NUMBER: PCT/AU01/00822
 ; PRIOR FILING DATE: 2001-07-06
 ; NUMBER OF SEQ ID NOS: 39
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 5
 ; LENGTH: 341
 ; TYPE: PRT
 ; ORGANISM: Haemophilus paragallinarum
 US-10-336-840-5

Query Match 17.9%; Score 162; DB 12; Length 341;
 Best Local Similarity 34.1%; Pred. No. 1.9e-07;

Matches 45; Conservative 24; Mismatches 53; Indels 10; Gaps 4;
 QY 56 EPAPVVEKPAVESGTTIASIYFDPDKYIKESQDQETLDEIV-KQAKENHMOVLLEGNTD 114
 DB 206 QSPAPVVEKPAVAKTFAALNSDVTFAFGKANLRPEAQNVLDGIYGEIAQLKSVQVLDAGYTD 265
 QY 115 EFGSSEYNOALGVKRTLSVKNALVIKGVKMDIKTISFGETKP-----KC-AQKTR- 165
 DB 266 RIGSEAYNQSLSEKRAQSVVDYLVAKGIPAGKISARGMGSENPVTGNTCDNVKARA 325
 QY 166 CYKENRRVDVKL 177
 DB 326 CLADDERVEISV 337

RESULT 4
 US-10-336-840-8
 ; Sequence 8, Application US/10336840
 ; Publication No. US20030219454A1
 ; GENERAL INFORMATION:
 ; APPLICANT: TERRY, TAMSIN DEBORAH
 ; APPLICANT: TSENG, HSING-JU
 ; APPLICANT: HOBBS, RHONDA IVY
 ; APPLICANT: JENNINGS, MICHAEL PAUL
 ; APPLICANT: DOWNES, JOHN
 ; TITLE OF INVENTION: HAEMAGGLUTININ ANTIGEN
 ; FILE REFERENCE: 37955-0007
 ; CURRENT APPLICATION NUMBER: US/10/336,840
 ; CURRENT FILING DATE: 2003-01-06
 ; PRIOR APPLICATION NUMBER: PCT/AU01/00822
 ; PRIOR FILING DATE: 2001-07-06
 ; NUMBER OF SEQ ID NOS: 39
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 8
 ; LENGTH: 341
 ; TYPE: PRT
 ; ORGANISM: Haemophilus paragallinarum
 US-10-336-840-8

Query Match 17.9%; Score 162; DB 12; Length 341;
 Best Local Similarity 34.1%; Pred. No. 1.9e-07;
 Matches 45; Conservative 24; Mismatches 53; Indels 10; Gaps 4;
 QY 56 EPAPVVEKPAVESGTTIASIYFDPDKYIKESQDQETLDEIV-KQAKENHMOVLLEGNTD 114
 DB 206 QSPAPVVEKPAVAKTFAALNSDVTFAFGKANLRPEAQNVLDGIYGEIAQLKSVQVLDAGYTD 265
 QY 115 EFGSSEYNOALGVKRTLSVKNALVIKGVKMDIKTISFGETKP-----KC-AQKTR- 165
 DB 266 RIGSEAYNQSLSEKRAQSVVDYLVAKGIPAGKISARGMGSENPVTGNTCDNVKARA 325
 QY 166 CYKENRRVDVKL 177
 DB 326 CLADDERVEISV 337

RESULT 5
 US-10-336-840-11
 ; Sequence 11, Application US/10336840
 ; Publication No. US20030219454A1
 ; GENERAL INFORMATION:
 ; APPLICANT: TERRY, TAMSIN DEBORAH
 ; APPLICANT: TSENG, HSING-JU
 ; APPLICANT: HOBBS, RHONDA IVY
 ; APPLICANT: JENNINGS, MICHAEL PAUL
 ; APPLICANT: DOWNES, JOHN
 ; TITLE OF INVENTION: HAEMAGGLUTININ ANTIGEN
 ; FILE REFERENCE: 37955-0007
 ; CURRENT APPLICATION NUMBER: US/10/336,840
 ; CURRENT FILING DATE: 2003-01-06
 ; PRIOR APPLICATION NUMBER: PCT/AU01/00822
 ; PRIOR FILING DATE: 2001-07-06
 ; NUMBER OF SEQ ID NOS: 39


```

RESULT 9
US-10-336-840-10
; Sequence 10, Application US/10336840
; Publication No. US20030219454A1
; GENERAL INFORMATION:
; APPLICANT: TERRY, TAMSIN DEBORAH
; APPLICANT: TSENG, HSING-JU
; APPLICANT: HOBBS, RHONDA IVY
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: DOWNES, JOHN
; TITLE OF INVENTION: HAEMAGGLUTININ ANTIGEN
; FILE REFERENCE: 37955-0007
; CURRENT APPLICATION NUMBER: US/10/336,840
; PRIOR FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: PCT/AU01/00822
; PRIOR FILING DATE: 2001-07-06
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 10
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Haemophilus paragallinarum
US-10-336-840-10

Query Match      17.9%; Score 162; DB 12; Length 344;
Best Local Similarity 34.1%; Pred. No. 1.9e-07;
Matches 45; Conservative 24; Mismatches 53; Indels 10; Gaps 4;

Qy 56 EPAPVVEKPAVESGTTIIASIIYDFDKYIKESDQETLDEIV-QKAKENHMOVLLEGNTD 114
Db 209 QSAFVPEKVVAKTFALNSDVTFAFGKANLRPEAQNVLDGIYGEIAQLKSVQVDLAGYTD 268

Qy 115 EFGSEYNQALGVKRTLSVKNALVIKGVKDMIKTISFGETKPKCAQKRECVKNNRVDVKL 177
Db 269 RIGSEAAANLKSQRADTVANLVSKGVAQEVISSTGYGEANPVTGAKCDVAKGRKALIA 328

Qy 166 CYKENRRVDVKL 177
Db 329 CLADREVEISV 340

RESULT 10
US-09-747-348-2
; Sequence 2, Application US/09747348
; Patent No. US20020123067A1
; GENERAL INFORMATION:
; APPLICANT: Andrew D. MURDIN, Raymond P. OOMEN, Joe WANG, Pamela DUNN
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there
; FILE REFERENCE: 77813-37
; CURRENT APPLICATION NUMBER: US/09/747,348
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 60/171,525
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 192
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-747-348-2

Query Match      17.7%; Score 160; DB 12; Length 192;
Best Local Similarity 33.3%; Pred. No. 1.3e-07;
Matches 40; Conservative 26; Mismatches 50; Indels 4; Gaps 3;

Qy 62 BEKPAVESGTTIIA--SIYDFDKYIK--ESDQETLDEIVQKAKENHMOVL--LEGNTD 114
Db 70 KEEKQYKSSQVAARNITFATDSYTIKGEENLAILTNLVHMKPKATLYIEGHTDERG 129

Qy 118 SSEYNQALGVKRTLSVKNALVIKGVKDMIKTISFGETKPKCAQKRECVKNNRVDVKL 177
Db 130 AASYNLALGARRANAIRKHEHLRKGISADRLSTISYKHEPLNSGHNELAWQNRTEFKI 189

Qy 166 CYKENRRVDVKL 177
Db 329 CLADREVEISV 340

RESULT 11
US-10-334-137-2
; Sequence 2, Application US/10334137
; Publication No. US20030161833A1
; GENERAL INFORMATION:
; APPLICANT: Andrew D. MURDIN, Raymond P. OOMEN, Joe WANG, Pamela DUNN
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there
; FILE REFERENCE: 77813-37
; CURRENT APPLICATION NUMBER: US/10/334,137
; PRIOR FILING DATE: 2002-12-31
; PRIOR APPLICATION NUMBER: US/09/747,348
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 60/171,525
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 192
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-334-137-2

Query Match      17.7%; Score 160; DB 12; Length 192;
Best Local Similarity 33.3%; Pred. No. 1.3e-07;
Matches 40; Conservative 26; Mismatches 50; Indels 4; Gaps 3;

Qy 62 BEKPAVESGTTIIA--SIYDFDKYIK--ESDQETLDEIVQKAKENHMOVL--LEGNTD 117
Db 70 KEEKQYKSSQVAARNITFATDSYTIKGEENLAILTNLVHMKPKATLYIEGHTDERG 129

Qy 118 SSEYNQALGVKRTLSVKNALVIKGVKDMIKTISFGETKPKCAQKRECVKNNRVDVKL 177
Db 130 AASYNLALGARRANAIRKHEHLRKGISADRLSTISYKHEPLNSGHNELAWQNRTEFKI 189

RESULT 12
US-10-336-840-3
; Sequence 3, Application US/10336840
; Publication No. US20030219454A1
; GENERAL INFORMATION:
; APPLICANT: TERRY, TAMSIN DEBORAH
; APPLICANT: TSENG, HSING-JU
; APPLICANT: HOBBS, RHONDA IVY
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: DOWNES, JOHN
; TITLE OF INVENTION: HAEMAGGLUTININ ANTIGEN
; FILE REFERENCE: 37955-0007
; CURRENT APPLICATION NUMBER: US/10/336,840
; PRIOR FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: PCT/AU01/00822
; PRIOR FILING DATE: 2001-07-06
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Haemophilus paragallinarum
US-10-336-840-3

Query Match      17.1%; Score 155; DB 12; Length 344;
Best Local Similarity 32.6%; Pred. No. 9.1e-07;
Matches 43; Conservative 25; Mismatches 54; Indels 10; Gaps 4;

Qy 56 EPAPVVEKPAVESGTTIIASIIYDFDKYIKESDQETLDEIV-QKAKENHMOVLLEGNTD 114
Db 209 QSAFVPEKVVAKTFALNSDVTFAFGKANLRPEAQNVLDGIYGEIAQLKSVQVDLAGYTD 268

Qy 115 EFGSEYNQALGVKRTLSVKNALVIKGVKDMIKTISFGETKPKCAQKRECVKNNRVDVKL 177
Db 269 RIGSEAAANLKSQRADTVANLVSKGVAQEVISSTGYGEANPVTGAKCDVAKGRKALIA 328

```

```

RESULT 11
US-10-334-137-2
; Sequence 2, Application US/10334137
; Publication No. US20030161833A1
; GENERAL INFORMATION:
; APPLICANT: Andrew D. MURDIN, Raymond P. OOMEN, Joe WANG, Pamela DUNN
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there
; FILE REFERENCE: 77813-37
; CURRENT APPLICATION NUMBER: US/10/334,137
; PRIOR FILING DATE: 2002-12-31
; PRIOR APPLICATION NUMBER: US/09/747,348
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 60/171,525
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 192
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-334-137-2

Query Match      17.7%; Score 160; DB 12; Length 192;
Best Local Similarity 33.3%; Pred. No. 1.3e-07;
Matches 40; Conservative 26; Mismatches 50; Indels 4; Gaps 3;

Qy 62 BEKPAVESGTTIIA--SIYDFDKYIK--ESDQETLDEIVQKAKENHMOVL--LEGNTD 117
Db 70 KEEKQYKSSQVAARNITFATDSYTIKGEENLAILTNLVHMKPKATLYIEGHTDERG 129

Qy 118 SSEYNQALGVKRTLSVKNALVIKGVKDMIKTISFGETKPKCAQKRECVKNNRVDVKL 177
Db 130 AASYNLALGARRANAIRKHEHLRKGISADRLSTISYKHEPLNSGHNELAWQNRTEFKI 189

RESULT 12
US-10-336-840-3
; Sequence 3, Application US/10336840
; Publication No. US20030219454A1
; GENERAL INFORMATION:
; APPLICANT: TERRY, TAMSIN DEBORAH
; APPLICANT: TSENG, HSING-JU
; APPLICANT: HOBBS, RHONDA IVY
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: DOWNES, JOHN
; TITLE OF INVENTION: HAEMAGGLUTININ ANTIGEN
; FILE REFERENCE: 37955-0007
; CURRENT APPLICATION NUMBER: US/10/336,840
; PRIOR FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: PCT/AU01/00822
; PRIOR FILING DATE: 2001-07-06
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Haemophilus paragallinarum
US-10-336-840-3

Query Match      17.1%; Score 155; DB 12; Length 344;
Best Local Similarity 32.6%; Pred. No. 9.1e-07;
Matches 43; Conservative 25; Mismatches 54; Indels 10; Gaps 4;

Qy 56 EPAPVVEKPAVESGTTIIASIIYDFDKYIKESDQETLDEIV-QKAKENHMOVLLEGNTD 114
Db 209 QSAFVPEKVVAKTFALNSDVTFAFGKANLRPEAQNVLDGIYGEIAQLKSVQVDLAGYTD 268

Qy 115 EFGSEYNQALGVKRTLSVKNALVIKGVKDMIKTISFGETKPKCAQKRECVKNNRVDVKL 177
Db 269 RIGSEAAANLKSQRADTVANLVSKGVAQEVISSTGYGEANPVTGAKCDVAKGRKALIA 328

```

```
QY 166 CYKENRRVDVKL 177
Db 329 CLADRRVEISV 340

RESULT 13
US-10-336-840-6
; Sequence 6, Application US/10336840
; Publication No. US20030219454A1
; GENERAL INFORMATION:
; APPLICANT: TERRY, TAMSIN DEBORAH
; APPLICANT: TSENG, HSING-JU
; APPLICANT: HOBBS, RHONDA IVY
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: DOWNES, JOHN
; TITLE OF INVENTION: HAEMAGGLUTININ ANTIGEN
; FILE REFERENCE: 37955-0007
; CURRENT APPLICATION NUMBER: US/10/336,840
; PRIOR FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: PCT/AU01/00822
; PRIOR FILING DATE: 2001-07-06
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Haemophilus paragallinarum
US-10-336-840-6

Query Match 17.1%; Score 155; DB 12; Length 344;
Best Local Similarity 32.6%; Pred. No. 9.1e-07;
Matches 43; Conservative 25; Mismatches 54; Indels 10; Gaps 4;

QY 56 EPAPVVEKPAVESGTTIASIYFDKYEIKESDQETLDEIV-OKAKENHMOVLLEGNTD 114
Db 209 QSAFVVEPKVAKTFALNSDVTFAFGKANLRPEAQNVLDGIYGEIAQLKSVQVDVAGYTD 268

QY 115 EFGSSEYNQALGVKRTLSVKNALVIKGVKDMIKTISFGETKP----KC-AQKTR----- 165
Db 269 RIGSEANLKLQRRADTVANLYLVSKGVAQEVISSTGYGEANPVTGAKCDTVKGRKALIA 328

QY 166 CYKENRRVDVKL 177
Db 329 CLADRRVEISV 340

RESULT 14
US-10-336-840-7
; Sequence 7, Application US/10336840
; Publication No. US20030219454A1
; GENERAL INFORMATION:
; APPLICANT: TERRY, TAMSIN DEBORAH
; APPLICANT: TSENG, HSING-JU
; APPLICANT: HOBBS, RHONDA IVY
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: DOWNES, JOHN
; TITLE OF INVENTION: HAEMAGGLUTININ ANTIGEN
; FILE REFERENCE: 37955-0007
; CURRENT APPLICATION NUMBER: US/10/336,840
; PRIOR FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: PCT/AU01/00822
; PRIOR FILING DATE: 2001-07-06
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 7
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Haemophilus paragallinarum
US-10-336-840-7

Query Match 17.1%; Score 155; DB 12; Length 344;
Best Local Similarity 32.6%; Pred. No. 9.1e-07;
Matches 43; Conservative 25; Mismatches 54; Indels 10; Gaps 4;

QY 56 EPAPVVEKPAVESGTTIASIYFDKYEIKESDQETLDEIV-OKAKENHMOVLLEGNTD 114
Db 209 QSAFVVEPKVAKTFALNSDVTFAFGKANLRPEAQNVLDGIYGEIAQLKSVQVDVAGYTD 268

QY 115 EFGSSEYNQALGVKRTLSVKNALVIKGVKDMIKTISFGETKP----KC-AQKTR----- 165
Db 269 RIGSEANLKLQRRADTVANLYLVSKGVAQEVISSTGYGEANPVTGAKCDTVKGRKALIA 328

QY 166 CYKENRRVDVKL 177
Db 329 CLADRRVEISV 340

RESULT 15
US-10-336-840-9
; Sequence 9, Application US/10336840
; Publication No. US20030219454A1
; GENERAL INFORMATION:
; APPLICANT: TERRY, TAMSIN DEBORAH
; APPLICANT: TSENG, HSING-JU
; APPLICANT: HOBBS, RHONDA IVY
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: DOWNES, JOHN
; TITLE OF INVENTION: HAEMAGGLUTININ ANTIGEN
; FILE REFERENCE: 37955-0007
; CURRENT APPLICATION NUMBER: US/10/336,840
; PRIOR FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: PCT/AU01/00822
; PRIOR FILING DATE: 2001-07-06
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 9
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Haemophilus paragallinarum
US-10-336-840-9

Query Match 17.1%; Score 155; DB 12; Length 344;
Best Local Similarity 32.6%; Pred. No. 9.1e-07;
Matches 43; Conservative 25; Mismatches 54; Indels 10; Gaps 4;

QY 56 EPAPVVEKPAVESGTTIASIYFDKYEIKESDQETLDEIV-OKAKENHMOVLLEGNTD 114
Db 209 QSAFVVEPKVAKTFALNSDVTFAFGKANLRPEAQNVLDGIYGEIAQLKSVQVDVAGYTD 268

QY 115 EFGSSEYNQALGVKRTLSVKNALVIKGVKDMIKTISFGETKP----KC-AQKTR----- 165
Db 269 RIGSEANLKLQRRADTVANLYLVSKGVAQEVISSTGYGEANPVTGAKCDTVKGRKALIA 328

QY 166 CYKENRRVDVKL 177
Db 329 CLADRRVEISV 340

Search completed: December 10, 2003, 18:40:31
Job time : 23.0236 secs
```



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:30:22 ; Search time 11.949 Seconds
(without alignments)
1440.645 Million cell updates/sec

Title: US-10-080-113-2
Perfect score: 906
Sequence: 1 MKRSSVFSFLVAFLLVAGCS.....AQKTRCYKENRRVDVKLMK 179

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	906	100.0	179	E64660	peptidoglycan asso
2	877	96.8	179	F71854	probable outer mem
3	250	27.6	165	I40769	Peptidoglycan asso
4	223	25.3	194	A45275	21k outer membrane
5	222.5	24.6	186	E82625	outer membrane pro
6	209.5	23.1	319	C81294	outer membrane fib
7	208	23.0	240	A37004	outer membrane cla
8	208	23.0	242	A81782	outer membrane pro
9	208	23.0	242	C81205	outer membrane pro
10	205	22.6	166	S53308	pall protein - Pse
11	198	21.9	188	C87649	OmpA family protei
12	194.5	21.5	236	A27894	outer membrane pro
13	193	21.3	168	G83525	outer membrane pro
14	191	21.1	210	A83516	probable outer mem
15	187.5	20.7	167	S58217	outer membrane pro
16	183	20.2	153	A28543	peptidoglycan-asso
17	178	19.6	168	AE0138	peptidoglycan-asso
18	177	19.5	155	F71637	peptidoglycan-asso
19	177	19.5	177	AE3013	omp16 protein [imp
20	177	19.5	177	B98271	omp16 protein [imp
21	176	19.4	168	I40346	omp16 protein - Br
22	176	19.4	168	AF3294	peptidoglycan-asso
23	176	19.4	202	G81553	peptidoglycan asso
24	174.5	19.3	238	I40703	outer membrane pro
25	174.5	19.3	359	S07298	outer membrane pro
26	172.5	19.0	173	1 LPECFG	peptidoglycan-asso
27	172.5	19.0	173	A85577	peptidoglycan-asso
28	172	19.0	173	H90725	peptidoglycan-asso
29	171.5	18.9	350	AI0626	outer membrane pro

ALIGNMENTS

RESULT 1

E64660
peptidoglycan associated lipoprotein precursor - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: E64660
R:Tomb, J.F.; White, O.; Kervatage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khatlani, H.G.; Glodok, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: E64660
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-179 <TOM>
A:Cross-references: GB:AE000619; GB:AE000511; NID:g2314276; PIDN:AAD08169.1; PID:g2314276
C:Superfamily: outer membrane protein A

Query Match 100.0%; Score 906; DB 2; Length 179;
Best Local Similarity 100.0%; Pred. No. 3.7e-63;
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKRSSVFSFLVAFLLVAGCSHKMDNKTIVAGDSVSAKTIVTAPVTTEPAPEKEEPKQEPAPV 60
DB 1 MKRSSVFSFLVAFLLVAGCSHKMDNKTIVAGDSVSAKTIVTAPVTTEPAPEKEEPKQEPAPV 60
QY 61 VEERPAVESGTTIIASIVYDFDKYIEKSDQETLDEIVQKAKENHMQVLLGNTDFGSGSE 120
DB 61 VEERPAVESGTTIIASIVYDFDKYIEKSDQETLDEIVQKAKENHMQVLLGNTDFGSGSE 120
QY 121 YNQALGVKRTLSVKNALVIGVEXDMKTISFGTETPKCAQKTRCYKENRRVDVKLMK 179
DB 121 YNQALGVKRTLSVKNALVIGVEXDMKTISFGTETPKCAQKTRCYKENRRVDVKLMK 179

RESULT 2

F71854
probable outer membrane protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: F71854
R:Alm, R.A.; King, L.S.L.; McI, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori.
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: F71854
A:Status: preliminary
A:Molecule type: DNA

A;Residues: 1-179 <ARN>
A;Cross-references: GB:AE001533; GB:AE001439; NID:94155636; PIDN:AAD06633.1; PID:9415564
A;Experimental source: strain 099
C;Genetics:
A;Gene: jhp1054
C;Superfamily: outer membrane protein A

Query Match 96.8%; Score 877; DB 2; Length 179;
Best Local Similarity 95.5%; Pred. No. 6.5e-61;
Matches 171; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 MKRSSVFLVAFLLVAGCSHKMDNKTIVAGDVSAKTQTAPVTPTEPAPEKEEPKQSPAPV 60
DB 1 MKRSSVFLVAFLLVAGCSHKMDNKTIVAGDVSAKTQTAPVTPTEPAPEKEEPKQSPAPV 60

QY 61 VEKPAVESGTHIIASIVDFDPKYEIKESDQETLDEIVQKAKENHMQVLLGNTDFGSS 120
DB 61 VEKPAVESGTHIIASIVDFDPKYEIKESDQETLDEIVQKAKENHMQVLLGNTDFGSS 120

QY 121 YNQALGVKRTLSVKNALVIGVEKMDIKTISFGETKPKCAQKRECYKKNRRVDVKLMK 179
DB 121 YNQALGVKRTLSVKNALVIGVEKMDIKTISFGESKPKCVQKTRCYKKNRRVDVKLMK 179

RESULT 3
140769
peptidoglycan associated lipoprotein precursor (omp18) Cj0113 [imported] - Campylobacter
C;Species: Campylobacter jejuni
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 03-Jun-2002
C;Accession: 140769; A81428
R;Burnens, A.; Stucki, U.; Nicolet, J.; Frey, J.
J. Clin. Microbiol. 33, 2826-2832, 1995
A;Title: Identification and characterization of an immunogenic outer membrane protein of
A;Reference number: 140769; MUID:9612122; PMID:8576327
A;Accession: 140769
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-165 <RSS>
A;Cross-references: EMBL:X83374; NID:91063273; PIDN:CAA58288.1; PID:91063274
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Bartel
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: A81428
A;Molecule type: DNA
A;Residues: 1-165 <PAR>
A;Cross-references: GB:AL139074; GB:AL111168; NID:96967505; PIDN:CAB72597.1; PID:9696760
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: omp18; pal; Cj0113
C;Superfamily: outer membrane protein A

Query Match 27.6%; Score 250; DB 2; Length 165;
Best Local Similarity 32.6%; Pred. No. 2.1e-12;
Matches 59; Conservative 36; Mismatches 68; Indels 18; Gaps 4;

QY 1 MKRSSVFLVAFLLVAGCSHKMDNKTIVAGDVSAKTQTAPVTPTEPAPEKEEPKQSPAPV 60
DB 1 MKRLFTSLAALVAVGCGCTK--STSVSGDSSVDS-----NRSGGSDGWD 45

QY 61 VEKPAVESGTHIIASIVDFDPKYEIKESDQETL--DEIVQKAKENHMQVLLGNTDFGSS 118
DB 46 IDSKISQLNLT-LNKVYDFDKENIRPDQMNVVSTNANFNTEVSGVSTVEGNCDEWGT 104

QY 119 SEYNQALGVKRTLSVKNALVIGVEKMDIKTISFGETKPKCAQKRECYKKNRRVDVKLM 178
DB 105 DEYNQALGLKRAKAVKEALIAKGNADRIAVKSYGTEPNVCTEKTACDAQNRRAEFKLS 164

QY 179 K 179
DB 165 R 165

RESULT 4

A45275

21X outer membrane protein OmpA - Bordetella avium

C;Species: Bordetella avium

C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 24-Nov-1999

C;Accession: A45275

R;Geny-Weeks, C.R.; Hultsch, A.L.; Kelly, S.M.; Keith, J.M.; Curtiss III, R.

J. Bacteriol. 174, 7729-7742, 1992

A;Title: Cloning and sequencing of a gene encoding a 21-kilodalton outer membrane protein

A;Reference number: A45275; MUID:93077456; PMID:1447140

A;Contents: 197

A;Accession: A45275

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: nucleic acid

A;Residues: 1-194 <GEN>

A;Cross-references: GB:M96550; NID:9144051; PIDN:AAA22979.1; PID:9144052

A;Note: sequence extracted from NCBI backbone (NCBI:P:119132)

C;Superfamily: outer membrane protein A

Query Match 25.3%; Score 229; DB 2; Length 194;
Best Local Similarity 28.6%; Pred. No. 1e-10;
Matches 59; Conservative 35; Mismatches 68; Indels 44; Gaps 5;

QY 1 MKRSSVFLVAFLLVAGCSHKMDNKTIVAGDVSAKTQT----- 39

DB 1 MNKPSKFLALAFRAV-----TASGVASQTVDNWRNPYGNVWKNGTNLCWRDA 50

QY 40 --APVTTEP-----APEKEEPKQEPAPVVEKPAVESGTHIIASIVDFDPKYEIKESDQETL 93

DB 51 FWTPTATGPGCDGVPVQAQPKPKAPMAAKV-----VFNAATDFDPKSTLKEPGRQLL 104

QY 94 DEIVQKAKENHMQVLL--GNTDFGSSBYNQALGVKRTLSVKNALVIGVEKMDIKTISF 152

DB 105 DOVAQAARAIDELIIIVGNTDTSIGTEAYNKLSERAASVKALVSKGIDPNRIYTEGK 164

QY 153 GETKPKCAQKRECYKKNRRVDVKLM 178

DB 165 GKLNPASNKTAEGRRNRVEIIV 190

RESULT 5

E82625

outer membrane protein P6 precursor XF1896 [imported] - Xylella fastidiosa (strain 9a5c)

C;Species: Xylella fastidiosa

C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C;Accession: E82625

R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A;Reference number: A82515; MUID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: E82625

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-186 <STM>

A;Cross-references: GB:AE004009; GB:AE003849; NID:99106980; PIDN:AAF84702.1; GSPDB:GN001

A;Experimental source: strain 9a5c

R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; B

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H

as-Neto, E.; Docena, C.; El-Dotry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krueger, J.E.; Kuramae, E.E.; Laig

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A;Reference number: A59328

A;Contents: annotation

C;Genetics:

A;Status: preliminary; not compared with conceptual translation

R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; E

A;Status: preliminary; not compared with conceptual translation

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
 ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
 Science 287, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
 A>Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
 A:Reference number: A81000; MUID:20175755; PMID:10710307
 A:Accession: C81205
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-242 <TET>
 A:Cross-references: GB:AE002394; GB:AE002098; NID:G7225599; PIDN:AAF40822.1; PID:G722560
 A:Experimental source: serogroup B, strain MC58
 C:Genetics:
 A:Gene: NMB0382

Query Match 23.08; Score 208; DB 2; Length 242;
 Best Local Similarity 29.08; Pred. No. 5.7e-09;
 Matches 64; Conservative 29; Mismatches 84; Indels 44; Gaps 6;
 Qy 1 MKRSSVFLVAFLLVAGCSHKDKNTVAGD-----VSATVQTA 40
 Db 5 LKLSALFVALLASGTAVAGEASVQYTVSGOSNEIVANNYGECKNAYFDKASQGRVECG 64
 Qy 41 PVTEPAPEKE-EPKQEPAPVVEKPAV--ESGTTIASIYDFDKYKESDQETLDEIV 97
 Db 65 DAVAAPPEPEPEPEPAPVAVVQAPQYVDETISLSAKTLFGFDKSLRAEAQDNLKVLA 124
 Qy 98 QKAKENMQ-VLLEGNTEFGSSSEYNQALGVKRTLSVKNALVKGVEKMIKITSFGETK 156
 Db 125 QRLSRTNVQSVRVGHTDFNGSKYQALSERAYVANNLVNGVPPVSRISAVGLGESQ 184
 Qy 157 PKCAQ-----KTRP-----CYKENRRVDVKL 177
 Db 185 AQMTQVCEAEVAKLGAKVSKAKREALIACIEPDRRDVKI 225

RESULT 10
 552308
 Ball protein - *Pseudomonas putida*
 C:Species: *Pseudomonas putida*
 C:Date: 08-May-1995 #sequence_revision 21-Jul-1995 #text_change 08-Oct-1999
 A:Accession: S52308
 R:Ramos-Gonzalez, M.I.; Ramos, J.
 submitted to the EMBL Data Library, July 1993
 A:Reference number: S52308
 A:Accession: S52308
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-166 <RAM>
 A:Cross-references: EMBL:X74218; NID:9861087; PIDN:CAA52294.1; PID:G663068
 C:Superfamily: outer membrane protein A

Query Match 22.68; Score 205; DB 2; Length 166;
 Best Local Similarity 29.48; Pred. No. 6.3e-09;
 Matches 52; Conservative 36; Mismatches 73; Indels 16; Gaps 4;
 Qy 1 MKRSSVFLVAFLLVAGCSHK-MDNKTAVGDSVSAKTQVAPVTEPAPEKEPKQEPAP 59
 Db 4 LKSGKFAALALAVAVGCSKSGDN---AGEAAVDPNRGYANTGAVDGS-----52
 Qy 60 VVEKPAVSGTTIASIYDFDKYKESDQETLDEIVQAKENHMQVLLEGNTEFGSS 119
 Db 53 -LSEEAALRA--ITTFYFEYDSDLKPEARALDVHAKDLKANGRVVLEGNTERGTR 108
 Qy 120 EYKQALGVKRTLSVKNALVKGVEKMIKITSFGETKPKCAQKTRCYKENRRVDVK 176
 Db 109 EYNVALGERRAKAVQRYLVLGVSQAQLELVSYGERPVAATGNDQSNQNRVLEUR 165

RESULT 11
 C87649
 OmpA family protein [imported] - *Caulobacter crescentus*
 C:Species: *Caulobacter crescentus*

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001.
 A:Accession: C87649
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A>Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A:Reference number: A87249; MUID:21173698; PMID:11259647
 A:Accession: C87649
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-188 <STO>
 A:Cross-references: GB:AB005673; NID:G13424913; PIDN:AAK25191.1; GSPDB:GN00149
 C:Genetics:
 A:Gene: CC3229

Query Match 21.98; Score 198; DB 2; Length 188;
 Best Local Similarity 34.98; Pred. No. 2.5e-08;
 Matches 51; Conservative 24; Mismatches 51; Indels 20; Gaps 4;
 Qy 41 PVTEPA---BEKEEPKQEPAPVVEKPAVESGTTIAS-----IYDFDKYK 86
 Db 37 PVTEPAQPPQPTPTPTTPAP-----PPVTSGILPGSVQDFVNVIGDRVVDTEVIR 91
 Qy 87 ESDQETLDEIVQ-KAKENHMQVLLEGNTEFGSSSEYNQALGVKRTLSVKNALVKGVEKD 145
 Db 92 ADAQPVLAGQVQLNRYSSVIRIEGNADERTREYNLALGARRANAVRDLAQGVSSA 151
 Qy 146 MIKITSFGETKPKCAQKTRCYKENR 171
 Db 152 RIETISFGKERPIDPGSEEAQAKNR 177

RESULT 12
 A27894
 outer membrane protein Pili precursor - *Neisseria gonorrhoeae*
 N:Alternate names: gonococcal protein III; Pili
 C:Species: *Neisseria gonorrhoeae*
 C:Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 08-Oct-1999
 R:Gotschlich, E.C.; Seiff, M.; Blake, M.S.
 J. Exp. Med. 165, 471-482, 1987
 A>Title: The DNA sequence of the structural gene of gonococcal protein III and the flank
 A:Reference number: A27894; MUID:87139801; PMID:3102671
 A:Accession: A27894
 A:Molecule type: DNA
 A:Residues: 1-236 <GOT>
 A:Cross-references: GB:X05105; NID:G44889; PIDN:CAA28752.1; PID:G44890
 R:Gotschlich, E.C.; Lyttton, E.J.; Seiff, M.
 Antonie Van Leeuwenhoek 53, 455-459, 1987
 A>Title: Gonococcal protein III. Purification and chemical characterization of the prote
 A:Reference number: A60817; MUID:88221126; PMID:3130783
 A:Accession: A60817
 A:Molecule type: protein
 A:Residues: 23-30, 'X', 32 <GO>
 C:Keywords: membrane protein
 F:1-32/Domain: signal sequence #status predicted <SIG>
 F:22-236/Product: Outer membrane protein Pili #status experimental <MAT>

Query Match 21.58; Score 194.5; DB 2; Length 236;
 Best Local Similarity 34.78; Pred. No. 6.2e-08;
 Matches 52; Conservative 23; Mismatches 50; Indels 25; Gaps 5;
 Qy 53 PKQEPAP--VVEKPAV--ESGTTIASIYDFDKYKESDQETLDEIVQAKENHMQ-V 107
 Db 70 PEPEPAPVAVVEQAPQYVDETISLSAKTLFGFDKSLRAEAQDNLKVLAQLSRTNVQSV 129
 Qy 108 LLEGNTEFGSSSEYNQALGVKRTLSVKNALVKGVEKMIKITSFGETKPKCAQ-----161
 Db 130 RVEGHTDFMGSEKYNQALSERAYVANNLVNGVPPASIRSAVGLGESQAQMTQVQCAEV 189
 Qy 162 -----KTRP-----CYKENRRVDVKL 177

Db 190 AKLGAKASKAKKREALIACIEPDRVDVKI 219

RESULT 13

G83525

outer membrane protein OprL precursor PA0973 [imported] - Pseudomonas aeruginosa (strain G83525)

C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C;Accession: G83525

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen

A;Reference number: A82950; PMID:20437337; PMID:10984043

A;Accession: G83525

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-168 <STO>

A;Cross-references: GB:AE004530; GB:AE004091; NID:g9946865; PIDN:AAG04362.1; GSPDB:GN001

A;Experimental source: strain PAO1

C;Genetics:

A;Gene: OprL; PA0973

C;Superfamily: outer membrane protein A

Query Match 21.3%; Score 193; DB 2; Length 168;

Best Local Similarity 27.9%; Pred. No. 5.4e-08;

Matches 51; Conservative 33; Mismatches 73; Indels 26; Gaps 4;

QY 1 MKRSSVFLVAFLLVAGSGHNDKTVAGDVSAKTQTAPVTTEPAPKEEPKQEPAPV 60

Db 4 LKFGKFAALAMAVAVGCCSK-----GGDASG-----EGANGVDPN---AGY 44

QY 61 VEKPAVESG-----TIIASIFDFDKYEIKESDQETLDEIVQKAKENHMOVLLEGT 113

Db 45 GANSQVAGDGLSDEALRAITTFYFSDSLKPEAMRALDVHAKDLKSGQRVVLGHT 104

QY 114 DFGSSEYNQALGVKRTLSVKNALVIKGVKMDIKTISFGETKPKCAQKTRCYKENRRV 173

Db 105 DERGTREYNALGERAKAVQRYLVLGQVSPAQLVELVSYGKERPVATGHDEQSWAQNRRV 164

QY 174 DVK 176

Db 165 ELK 167

RESULT 14

A83516

probable outer membrane protein PA1041 [imported] - Pseudomonas aeruginosa (strain PAO1)

C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C;Accession: A83516

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen

A;Reference number: A82950; PMID:20437337; PMID:10984043

A;Accession: A83516

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-210 <STO>

A;Cross-references: GB:AE004536; GB:AE004091; NID:g9946945; PIDN:AAG04430.1; GSPDB:GN001

A;Experimental source: strain PAO1

C;Genetics:

A;Gene: PA1041

Query Match 21.1%; Score 191; DB 2; Length 210;

Best Local Similarity 33.1%; Pred. No. 1e-07;

Matches 48; Conservative 27; Mismatches 58; Indels 12; Gaps 4;

QY 38 QTAPVTTEPAPKEEPKQEPAPVVEKPAVESGTTIIASIFDFDKYEIKESDQETLDEIV 97

Db 74 QVAP-----PPQPVEEVAPPFPVVKETIV-----VRDLHFAFDSSKVDAAADSEKLNIA 124

QY 98 QKAKEN--HMQVLLEGNTEFGSSSEYNQALGVKRTLSVKNALVIKGVKMDIKTISFGET 154

Db 125 ERLGGEAASLTSLTGHTDSVGSDAYNQKLSERRANAVANVLIDAGVPSSIIIVGQGLGE 184

QY 155 TKPKCAQKTRCYKENRRVDVKLMK 179

Db 185 SQPVADNKTREGRAENRRVEILIKR 209

RESULT 15

S58217

outer membrane protein - Pseudomonas aeruginosa

C;Species: Pseudomonas aeruginosa

C;Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 26-Aug-1999

C;Accession: S58217

R;Lim, A.; de Vos, D.; Brauns, M.; Gaballa, A.; Hamers, R.; Cornelis, P.

submitted to the EMBL Data Library, August 1995

A;Description: Molecular cloning, nucleotide sequence, and characterization of OprL, a

A;Reference number: S58216

A;Accession: S58217

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-167 <LIM>

A;Cross-references: EMBL:Z50191

C;Superfamily: outer membrane protein A

Query Match 20.7%; Score 187.5; DB 2; Length 167;

Best Local Similarity 27.9%; Pred. No. 1.4e-07;

Matches 51; Conservative 34; Mismatches 71; Indels 27; Gaps 5;

QY 1 MKRSSVFLVAFLLVAGSGHNDKTVAGDVSAKTQTAPVTTEPAPKEEPKQEPAPV 60

Db 4 LKFGKFAALAMAVAVGCCSK-----GGDASG-----EGANGVDPN---AGY 44

QY 61 VEKPAVESG-----TIIASIFDFDKYEIKESDQETLDEIVQKAKENHMOVLLEGT 113

Db 45 GANSQVAGDGLSDEALRAITTFYFSDSLKPEAMRALDVHAKDLKSGQRVVLGHT 104

QY 114 DFGSSEYNQALGVKRTLSVKNALVIKGVKMDIKTISFGETKPKCAQKTRCYKENRRV 173

Db 105 DERGTREYNALGERAKAVQRYLVLGQVSR-QLELVSYGKERPVATGHDEQSWAQNRRV 163

QY 174 DVK 176

Db 164 ELK 166

Search completed: December 10, 2003, 18:37:40

Job time : 12.949 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:29:11 ; Search time 7.0288 Seconds
(without alignments)
1197.613 Million cell updates/sec

Title: US-10-080-113-2
Perfect score: 906
Sequence: 1 MKRSSVFSFLVAFLLVAGCS.....AOKTRECYKENRRVDVVKLMK 179

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	229	25.3	194	1 OMPA_BORAV	Q05146 bordetella
2	208	23.0	242	1 OMP4_NEIMA	P38367 neisseria m
3	205	22.6	166	1 PAL_PSEPK	P43036 pseudomonas
4	194.5	21.5	236	1 OMP3_NEIGO	P07050 neisseria g
5	190	21.0	176	1 OMP6_RHIME	Q926c3 rhizobium m
6	185	20.4	168	1 OMP6_RHIL0	Q98f85 rhizobium l
7	183	20.2	153	1 PAL_HABIN	P10324 haemophilus
8	180.5	19.9	150	1 PAL_PASMU	Q51886 pasteurilla
9	177	19.5	177	1 OMP16_AGR75	Q89155 agrobacteri
10	176	19.4	168	1 OMP16_BRUME	Q44662 brucella me
11	174.5	19.3	238	1 OMPA_CITFR	P24016 citrobacter
12	174.5	19.3	359	1 OMPA_SERMA	P04845 serraria ma
13	172.5	19.0	173	1 PAL_ECOLI	P01776 escherichia
14	171.5	18.9	350	1 OMPA_SALTY	P02936 salmonella
15	171	18.9	243	1 OMPA_ESCHE	P24754 escherichia
16	169	18.7	344	1 OMPA_KLEPN	P24017 klebsiella
17	164	18.1	350	1 OMPA_ENTAE	P09146 enterobacte
18	163	18.0	243	1 OMPA_ESCFE	P24747 escherichia
19	163	18.0	346	1 OMPA_ECOLI	P02934 escherichia
20	163	18.0	351	1 OMPA_SHIDY	P02935 shigella dy
21	158	17.4	326	1 PORF_PSEFL	P37726 pseudomonas
22	153.5	16.9	349	1 OMPA_BUCAL	P57414 buchnera ap
23	153	16.9	341	1 OMPA_ESCBL	Q99124 escherichia
24	153	16.9	350	1 PORF_PSEAE	P13794 pseudomonas
25	152	16.8	243	1 OMPA_SEROD	P24755 serraria od
26	146.5	16.2	344	1 POSF_PSESY	P22263 pseudomonas
27	145.5	16.1	353	1 OMS1_HABIN	P43840 haemophilus
28	141.5	15.6	353	1 OMS2_HABIN	P38368 haemophilus
29	138.5	15.3	176	1 PAL_LEGPN	P26493 legionella
30	135.5	15.0	359	1 OMS3_HABIN	P45996 haemophilus
31	128	14.1	347	1 OMS4_BUCAP	Q8x914 buchnera ap
32	127	14.0	326	1 Y899_WYCTUP	Q10557 mycobacteri
33	121.5	13.4	219	1 YIAD_ECOLI	P37665 escherichia

34	113	12.5	160	1 YFIB_ECOLI	P07021 escherichia
35	108	11.9	261	1 MOTB_BACSU	P28612 bacillus su
36	103	11.4	242	1 Y376_UREPA	C99pb5 ureaplasma
37	100.5	11.1	242	1 YTXE_BACSU	P39064 bacillus su
38	97	10.7	391	1 OMA1_PORGI	Q983x8 porphyromon
39	93.5	10.3	238	1 MOTB_TREPA	C07857 treponema p
40	91	10.0	257	1 MOTB_HELPFJ	Q92159 helicobacte
41	90.5	10.0	351	1 YIIG_ECOLI	P32151 escherichia
42	89.5	9.9	380	1 OMA0_PORGI	Q983x8 porphyromon
43	89	9.8	257	1 TONB_HELPY	P56427 helicobacte
44	89	9.8	285	1 TONB_HELPY	C02589 helicobacte
45	88	9.7	1234	1 B3A2_RAT	P23347 rattus norv

ALIGNMENTS

RESULT 1

OMPA_BORAV STANDARD; PRT; 194 AA.

AC Q05146;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Outer membrane protein A precursor.
GN OMPA.
OS Bordetella avium.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=521;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=197;
RX MEDLINE=93077456; PubMed=1447140;
RA Gentry-Weeks C.R., Hultsch A.-L., Kelly S.M., Keith J.M.,
RA Curtis R. III;
RT "Cloning and sequencing of a gene encoding a 21-kilodalton outer
membrane protein from Bordetella avium and expression of the gene in
Salmonella typhimurium".
RT J. Bacteriol. 174:7729-7742(1992).
RL -!- FUNCTION: STRUCTURAL PROTEIN THAT MAY PROTECT THE INTEGRITY OF
THE BACTERIUM.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M96550; AAA22979.1; -;
CC PIR; A45275; A45275.
CC InterPro; IPR006664; Bac OmpA.
CC InterPro; IPR001035; MotY.
CC InterPro; IPR006665; OmpA/MotB.
CC InterPro; IPR006690; OMPA_LIKE.
CC Pfam; PF00691; OmpA; 1.
CC PRINTS; PR01023; NARFLMOTY.
CC PRINTS; PR01021; OMPADOMAIN.
CC ProDom; PD000930; OmpA/MotB; 1.
CC PROSITE; PS01068; OMPA; 1.
CC Outer membrane; Transmembrane; Porin; Antigen; Signal.
CC SIGNAL 1 24 POTENTIAL.
FT CHAIN 1 194 OUTER MEMBRANE PROTEIN A.
FT DOMAIN 25 76 PRO-RICH.
FT DOMAIN 121 165 OMPA-LIKE.
SQ SEQUENCE 194 AA; 21115 MW; 1A25F2A6367DAE7A CRC64;
Query Match 25.3%; Score 229; DB 1; Length 194;
Best Local Similarity 28.6%; Pred. No. 1.3e-11;

Matches 59; Conservative 35; Mismatches 68; Indels 44; Gaps 5;

QY 1 MKRSVSVFLVAFLLVAGCGHKMDNKTAVGDAKTVQT----- 39
 DB 1 MNKPSKFLALFAAV-----TASGVASQTVDMNRPNYGNWKNGTNELCWRDA 50
 QY 40 --APVTTEP---APEKEBPQBPAPVVEKPAVESGTIIASIVYDFDKYKESDQETL 93
 DB 51 FWTPATGIPGCDGVPAVQAQKPKAPPAKV-----VFNADTFDFDKSTLPEGRQLL 104
 QY 94 DEIVQKAKENHMQVLE-GNTDFGSGSEYNQALGVKRTLSVKNALVIKGVKMDKMTISF 152
 DB 105 DQVAQAARADLETTIIVAGTDSITGEYANKLSERRAASVKAYLVSKGIDPNRIYTEGK 164
 QY 153 GETPKPKCAQKTRCYCKENRRVDVKLM 178
 DB 165 GKLNPIASNKTAAGRARRRVEIIV 190

RESULT 2

OMP4_NEIMA STANDARD; PRT; 242 AA.

AC P39367;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein class 4 precursor.
 GN RMPM OR NMA2105 OR NMB0382.
 OS Neisseria meningitidis (serogroup A), and
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=65699, 491;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=CCUG 18241 / M986 / Serogroup B / Serotype 2;
 RX MEDLINE=9277523; PubMed=2499543;
 RA Klugman K.P., Gotschlich E.C., Blake M.S.;
 RT "Sequence of the structural gene (rmpM) for the class 4 outer
 RT membrane protein of *Neisseria meningitidis*, homology of the protein
 RT to gonococcal protein III and *Escherichia coli* OmpA, and construction
 RT of meningococcal strains that lack class 4 protein.";
 RL Infect. Immun. 57:2066-2071(1989).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holtroyd S.,
 RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrall B.G.;
 RT "Complete DNA sequence of a serogroup A strain of *Neisseria*
 RT meningitidis Z2491.";
 RL Nature 404:502-506(2000).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=MC58 / Serogroup B;
 RX MEDLINE=2015755; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Citron H., Clark E.B.,
 RA Cotton M.D., Uitterback T.K., Khouri H., Qin H., Vamathevan J.,
 RA Gill J., Scarlato V., Masignani V., Pizzo M., Grandi G., Sun L.,
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
 RT "Complete genome sequence of *Neisseria meningitidis* serogroup B strain
 RT MC58.";
 RL Science 287:1809-1815(2000).
 CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -I- SIMILARITY: BELONGS TO THE OMPA FAMILY. STRONG, TO N.GONORRHOEAE

P.III.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AL162758; CAB85320.1; -
 CC EMBL; AE002394; AAF40822.1; -
 CC PIR; A37004; A37004.
 DR PIR; A81782; A81782.
 DR PIR; C81205; C81205.
 DR TIGR; NMB0382; -
 DR InterPro; IPR006664; Bac OmpA.
 DR InterPro; IPR006665; OmpA/MCB.
 DR InterPro; IPR006690; OMPA_LIKE.
 DR Pfam; PF00691; OmpA; 1.
 DR PRINTS; PR01021; OMPADOMAIN.
 DR ProDom; PD000930; OmpA/MCB; 1.
 DR PROSITE; PS01068; OMPA; 1.
 DR Outer membrane; Porin; Transmembrane; Signal; Repeat;
 KW Complete proteome.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 242 OUTER MEMBRANE PROTEIN CLASS 4.
 FT DOMAIN 69 82 7 X 2 AA TANDEN REPEATS OF X-P.
 FT REPEAT 69 70 1.
 FT REPEAT 71 72 2.
 FT REPEAT 73 74 3.
 FT REPEAT 75 76 4.
 FT REPEAT 77 78 5.
 FT REPEAT 79 80 6.
 FT REPEAT 81 82 7.
 FT DOMAIN 137 181 OMPA-LIKE.
 FT DISULFID 191 214 BY SIMILARITY.
 FT VARIANT 78 129 MISSING (IN STRAIN CCUG 18241).
 FT VARIANT 128 129 GQ -> SR (IN STRAIN MC58).
 FT VARIANT 132 132 I -> V (IN STRAIN MC58).
 SQ SEQUENCE 242 AA; 26140 MW; 5CCAA490236B1D62 CRC64;

Query Match Similarity 23.0%; Score 208; DB 1; Length 242;

Best Local Similarity 23.0%; Pred. No. 8.1e-10;
 Matches 64; Conservative 30; Mismatches 83; Indels 44; Gaps 6;

QY 1 MKRSVSVFLVAFLLVAGCGHKMDNKTAVG-----VSATVQTA 40
 DB 5 LKLSALFVALLASCTAVAGEASVQGYTVSGSNEIVNNYGEWKNAYFDKASQGRVECG 64
 QY 41 PVTTPEAPEKE-EPKQBPAPVVEKPAV--ESGTIIASIVYDFDKYKESDQETLDEIV 97
 DB 65 DAVAAPPE 124
 QY 98 QKAKENHMQ-VLLEGTDFGSGSEYNQALGVKRTLSVKNALVIKGVKMDKMTISFQETK 156
 DB 125 QRLQGTNIQSVRVGHTDFMGSDKYNQALSERRAYVYVNNLVNNGVPSVRSIAGVLSG 184
 QY 157 PKCAQ-----KTR-----CYKENRRVDVKL 177
 DB 185 AQMTQVCEAEVAKLGAKVSKAKKREALIACIEPDRRDVVKI 225

RESULT 3

PAL_PSEBK STANDARD; PRT; 166 AA.
 ID PAL_PSEBK
 AC P43036;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Peptidoglycan-associated lipoprotein precursor.
 GN PAL OR PALI OR OPRL OR PP1223.
 OS Pseudomonas putida (strain KT2440), and

```

120 QY EYNQALGVKRTLSVKNALVIGVKQDKMIKTIISFGETPKCAQKTRCYKENRRVDVK 176
    ||| ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
109 DB EYNMALGERAKAVQRYLVLGVSQAQLSVSGEERPVATGNDQSWAQNRRVELR 165

RESULT 4
OMP3_NEIGO          STANDARD;          PRT;          236 AA.
ID      OMP3_NEIGO          F07050;
AC      F07050;
DT      01-APR-1988 (Rel. 07, Created)
DT      01-APR-1988 (Rel. 07, Last sequence update)
DT      01-FEB-1994 (Rel. 28, Last annotation update)
DE      Outer membrane protein F.III precursor (Gonococcal protein III)
DE      (PIII).
OS      Neisseria gonorrhoeae.
OC      Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC      Neisseriaceae; Neisseria.
OX      NCBI_TaxID=485;
RN      [1]_TaxID=485;
RP      SEQUENCE FROM N.A.
RP      MEDLINE=87139801; PubMed=3102671;
RA      Gotschlich E.C., Seiff M., Blake M.S.;
RA      "The DNA sequence of the structural gene of gonococcal protein III
RT      and the flanking region containing a repetitive sequence. Homology of
RT      protein III with enterobacterial OmpA proteins.";
RT      J. Exp. Med. 165:471-482(1987).
RL      J. EXP. MED. 165:471-482(1987).
CC      -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC      -!- MISCELLANEOUS: PIII IS CLOSELY ASSOCIATED WITH PART, BUT NOT ALL
CC      OF THE PI MOLECULES IN THE GONOCOCCAL OUTER MEMBRANE.
CC      -!- MISCELLANEOUS: A PORTION OF THE PIII IS EXPOSED TO THE SURFACE IN
CC      INTERACT GONOCOCCI; THE PROTEIN CAN BE LABELED WITH IODINE AND
CC      INTERACTS WITH WABS.
CC      -!- SIMILARITY: BELONGS TO THE OMPA FAMILY. STRONG, TO N.MENINGITIDIS
CC      RMPM.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
-----
EMBL; X05105; CAA28752.1; -.
PIR; A27894; A27894.
InterPro; IPR006664; Bac OmpA.
InterPro; IPR006665; OmpA/MctB.
InterPro; IPR006690; OMPA-LIKE.
Pfam; PF00691; OmpA; 1.
PRINTS; PR01021; OMPADOMAIN.
ProDom; PD000930; OmpA/MctB; 1.
PROSITE; PS01068; OMPA; 1.
Outer membrane; Forin; Transmembrane; Antigen; Signal; Repeat.
SIGNAL          1      22
FT CHAIN        23      236
FT DOMAIN       69      76
FT REPEAT       69      70
FT REPEAT       71      72
FT REPEAT       73      74
FT REPEAT       75      76
FT DOMAIN       131     175
FT DISULFID     185     208
FT SEQUENCE     236 AA; 25540 MW; DAEFAAECA66FB199 CRC64;

Query Match
Best Local Similarity 34.7%; Pred. No. 9.6e-09;
Matches 52; Conservative 23; Mismatches 50; Indels 25; Gaps 5;

53 QY PKQEPAP--VVEKPAV--ESGTLIASIYFDFKYEIKESDQETLDEIVQKAKENHQ-V 107
    | : ||| ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
70 DB PEPEPAPVAVPEAQPYVDETISLAKTLFGFDKDSLRAEACDNLKVLQAQRLSRITNVQSV 129

```

108 LLEGTDFGSEYNOALGVKRTLSVKNALVIGVEKDMIKTISFGETPKCAQ-----161
 130 RVEGTDPMGSEKYNQALSERAYVANNVLSVGPASRISAVLGESQAQMTQVQAEV 189
 162 -----KTRP-----CYKENRRVDVKL 177
 190 AKLGAKASKAKREALIACIEPDPRVDVKI 219

RESULT 5
 OM16_RHIME STANDARD; PRT; 176 AA.
 AC Q926C3;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Outer membrane lipoprotein omp16 homolog precursor.
 GN OMP16 OR PAL OR R02738 OR SMC02942.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
 RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ransperger U.,
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.,
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021".
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 CC -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE OMPA FAMILY. PAL SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AL591791; CAC47317.1; -
 DR InterPro; IPR006664; Bac_OmpA.
 DR InterPro; IPR006665; OmpA/MotB.
 DR Pfam; PF00691; OmpA; 1.
 DR PRINTS; PR01021; OMPADOMAIN.
 DR ProDom; PD000930; OmpA/MotB; 1.
 DR PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
 DR PROSITE; PS01068; OMPA; FALSE NEG.
 KW Membrane; Outer membrane; Lipoprotein; signal; Complete proteome.
 FT SIGNAL 1 32 POTENTIAL.
 FT CHAIN 33 176 OUTER MEMBRANE LIPOPROTEIN OMP16 HOMOLOG.
 FT LIPID 33 33 N-ACYL DIGLYCERIDE (POTENTIAL).
 FT DOMAIN 103 147 OMPA-LIKE.
 FT SEQUENCE 176 AA; 18735 MW; A6F162CB35042268 CRC64;
 Query Match 21.0%; Score 190; DB 1; Length 176;
 Best Local Similarity 31.0%; Pred. No. 1.5e-08;
 Matches 53; Conservative 27; Mismatches 69; Indels 22; Gaps 4;
 QY 3 RSVFSLVAFLLVAGCSHKMDKTV--AGDVSAKTQVTPVTPAPEKEPQEPAPVVE 62
 DB 17 RNPVMTALVMTALAGCASK---KNLPNDAAAGLGLGAGAAT-----PGSQ 59
 QY 63 EKPAVESGTTIATSYDFDKYEIKESDQETLDEIVQ--KAKENHMVQLLGNTDFGSSSEY 121
 DB 59 QDFVNVGD---RIFFDTDSISRAQAQLDRQAQWLAKYPNYGITIGHADERTREY 115

122 NOALGVKRTLSVKNALVIGVEKDMIKTISFGETPKCAQKTRCYKENRR 172
 116 NLALGARAAATRDYLSRGVFGNRMRTISYGEKPVAVCDDISCSQNR 166

RESULT 6
 OM16_RHILO STANDARD; PRT; 168 AA.
 AC Q98F5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Outer membrane lipoprotein omp16 homolog precursor.
 GN OMP16 OR MLL3887.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti".
 RL DNA Res. 7:331-338(2000).
 CC -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE OMPA FAMILY. PAL SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AP003003; BAB50682.1; -
 DR InterPro; IPR006664; Bac_OmpA.
 DR InterPro; IPR006665; OmpA/MotB.
 DR InterPro; IPR006690; OMPA_LIKE.
 DR Pfam; PF00691; OmpA; 1.
 DR PRINTS; PR01021; OMPADOMAIN.
 DR ProDom; PD000930; OmpA/MotB; 1.
 DR PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
 DR PROSITE; PS01068; OMPA; FALSE NEG.
 KW Membrane; Outer membrane; Lipoprotein; signal; Complete proteome.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 168 OUTER MEMBRANE LIPOPROTEIN OMP16 HOMOLOG.
 FT LIPID 25 25 N-ACYL DIGLYCERIDE (POTENTIAL).
 FT DOMAIN 95 139 OMPA-LIKE.
 FT SEQUENCE 168 AA; 17765 MW; 03E260D3BA1ECB3 CRC64;
 Query Match 20.4%; Score 185; DB 1; Length 168;
 Best Local Similarity 30.9%; Pred. No. 3.7e-08;
 Matches 54; Conservative 24; Mismatches 71; Indels 26; Gaps 5;
 QY 1 MKRSVFSFLVAFLLVAGCSHKMDKTV--AGDVSAKTQVTPVTPAPEKEPQEPAPVVE 58
 DB 7 LTRPVMIALVMTALAGCASK---KTPNNAADLGLGAGAA-----T 46
 QY 59 PVVEEKPAVESGTTIATSYDFDKYEIKESDQETLDEIVQKAKE--NMQVLLGNTDFG 117
 DB 47 PGSAQDETIVGID---RIFFDTDSISRAQAQLDRQAQWLAKYPNYGITIGHADERTREY 103
 QY 118 SSEYNQALGVKRTLSVKNALVIGVEKDMIKTISFGETPKCAQKTRCYKENRR 172

Db 104 TREYNALGARAAARDPLVSGVASSRLKTSYKGRPVAVCDDISCSQNR 158

RESULT 7

PAL_HAEN STANDARD; PRT; 153 AA.

AC P10324;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-MAR-1989 (Rel. 10, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Outer membrane protein P6 precursor (OMP P6) (15 kDa peptidoglycan-associated lipoprotein) (PC protein).

GN PAL OR OMP56 OR HI0381.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Haemophilus.

OX NCBI_TaxID=727;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RX MEDLINE=88115138; PubMed=2828309;

RA Deich R.A., Matcalif B.J., Finn C.W., Farley J.E., Green B.A.;

RT "Cloning of genes encoding a 15,000-dalton peptidoglycan-associated outer membrane lipoprotein and an antigenically related 15,000-dalton protein from Haemophilus influenzae.";

RL J. Bacteriol. 170:489-498(1988).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=88085463; PubMed=3257200;

RA Nelson M.B., Apicella M.A., Murphy T.F., Vankeulen H., Spotila L.D.,

RA Rekesh D.;

RT "Cloning and sequencing of Haemophilus influenzae outer membrane protein P6.";

RL Infect. Immun. 56:128-134(1988).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=Ed / KM20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,

RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,

RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,

RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,

RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,

RA Venter J.C.;

RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";

RL Science 269:496-512(1995).

CC -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid anchor.

CC -!- SIMILARITY: BELONGS TO THE OMPA FAMILY. PAL SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; M19391; AAA2494.1; -

CC EMBL; M18878; AAA2494.1; -

CC EMBL; U32722; AAC22039.1; -

CC PIR; A28543; A28543.

CC TIGR; HI0381; -

CC InterPro; IPR006664; Bac OmpA.

CC InterPro; IPR006665; OmpA/MotB.

CC InterPro; IPR006650; OMPA-LIKE.

CC Pfam; PF00691; OmpA; 1.

CC PRINTS; PR01021; OMPADOMAIN.

CC ProDom; PD000930; OmpA/MotB; 1.

CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.

DR PROSITE; PS01068; OMPA; 1.

KW Outer membrane; Signal; Lipoprotein; Complete proteome.

FT SIGNAL 19

FT CHAIN 20 153 OUTER MEMBRANE PROTEIN P6.

FT LIPID 20 20 N-ACYL DIGLYCERIDE.

FT DOMAIN 85 129 OMPA-LIKE.

SQ SEQUENCE 153 AA; 16108 MW; 3DF358122EE17A11 CRC64;

Query Match 20.28; Score 183; DB 1; Length 153;

Best Local Similarity 31.9%; Pred. No. 4.8e-08;

Matches 58; Conservative 17; Mismatches 49; Indels 58; Gaps 6;

Qy 11 VAFLLVAG-----CSHKMDNKTVDGVSAAKTVQAPVTEPAPEKEPKQSPAPVVEE 63

Db 5 VKSLVAGSVAAALACS--SSNDAAGGAAQFF----- 36

Qy 64 KPAVESGTIIA-----SIYFDFDYKEIKESQETILD-----EIVOKAKENHMVLE 110

Db 37 -----GGYSVADLQQRNTYVFGDKYDITGEYVQILDAAHAYLNATPAK-----VLVE 86

Qy 111 GNTDFGSSVNOALGVKRTLSVKNALVIKGVKDMIKTISFGETPKCAQKTRCYKEN 170

Db 87 GNTDERGTPTNYALGQRRADAVKGLAGKVDAGKLTGTVSYGEEKPAVLGHDEAYSKN 146

Qy 171 RR 172

Db 147 RR 148

RESULT 8

PAL_PASMU STANDARD; PRT; 150 AA.

AC Q51886;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Outer membrane protein P6 precursor (OMP P6) (P6-like) (Peptidoglycan-associated lipoprotein).

GN PAL OR PM0966.

OS Pasteurella multocida.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Pasteurella.

OX NCBI_TaxID=747;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=T16;

RX MEDLINE=95172751; PubMed=7868272;

RA Kasten R.W., Hansen L.M., Hinojoza J., Bieber D., Ruehl W.W.,

RA Hirsch D.C.;

RT "Pasteurella multocida produces a protein with homology to the P6 outer membrane protein of Haemophilus influenzae.";

RL Infect. Immun. 63:989-993(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Pm70.

RX MEDLINE=21145866; PubMed=11248100;

RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;

RT "Complete genomic sequence of Pasteurella multocida Pm70.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).

CC -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid anchor (By similarity).

CC -!- SIMILARITY: BELONGS TO THE OMPA FAMILY. PAL SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; U16849; AAA79373.1; -

CC EMBL; AE006136; AAA03050.1; -

```

DR InterPro; IPR006664; Bac OmpA.
DR InterPro; IPR006665; OmpA/MotB.
DR InterPro; IPR006690; OmpA LIKE.
DR Pfam; PF00691; OmpA; 1.
DR PRINTS; PR01021; OMPADOMAIN.
DR ProDom; PD000930; OmpA/MotB; 1.
DR PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
DR PROSITE; PS01068; OmpA; 1.
KW Outer membrane; Signal; Lipoprotein; Complete proteome.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 150 OUTER MEMBRANE PROTEIN P6.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (BY SIMILARITY).
FT DOMAIN 82 126 OMPA-LIKE.
SQ SEQUENCE 150 AA; 16213 MW; 170E7B13D2B9ED6C CRC64;

Query Match 19.98; Score 180.5; DB 1; Length 150;
Best Local Similarity 41.68; Pred. No. 7.4e-08;
Matches 42; Conservative 16; Mismatches 38; Indels 5; Gaps 2;

Oy 75 SIYDFDKYIKESDQSTDEIVQKAKEN---HMQVLEGTDFGSGSEYNOALGVKRTL 131
Db 47 TVYGFDPKYNIEGYVQLD--AAFLNATPATKVVVEGTDERTGTPVNIALGORRAD 104

Oy 132 SVKNALVIGKVEKDMIKTISFGETKPKCAQKTRCYKENR 172
Db 105 AVKHLSAKGVQAGQVSTVSYGSEKPAVLGHDEAAYSQNR 145

RESULT 9
OM16_AGR75 STANDARD; PRT; 177 AA.
AC Q8U9L5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Outer membrane lipoprotein omp16 homolog precursor.
GN OMP16 OR PALA OR AU03713 OR AGR_L_2246.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58."
RL Science 294:2317-2323 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Quorillo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Humiel X., Gordon J., Vaudin M., Lartchouk O., Epp A., Liu F.,
RA Wallam C., Allinger M., Daughy D., Scott C., Lappas C., Markelz B.,
RA Flegan C., Crowell C., Gursan J., Lomo C., Sear C., Strub G.,
RA Cleo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
Agrobacterium tumefaciens C58."
RL Science 294:2323-2328 (2001).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
anchor (By similarity).
CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY. PAL SUBFAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE009302; AAL44523.1; -.
DR EMBL; AE008312; AAK89692.1; -.
DR PIR; AE3013; AE3013.
DR PIR; B98271; B98271.
DR InterPro; IPR006664; Bac OmpA.
DR InterPro; IPR006665; OmpA/MotB.
DR InterPro; IPR006690; OmpA LIKE.
DR Pfam; PF00691; OmpA; 1.
DR PRINTS; PR01021; OMPADOMAIN.
DR ProDom; PD000930; OmpA/MotB; 1.
DR PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
DR PROSITE; PS01068; OmpA; FALSE NEG.
KW Membrane; Outer membrane; Lipoprotein; Signal; Complete proteome.
FT SIGNAL 1 32 POTENTIAL.
FT CHAIN 33 177 OUTER MEMBRANE LIPOPROTEIN OMP16 HOMOLOG.
FT LIPID 33 33 N-ACYL DIGLYCERIDE (POTENTIAL).
FT DOMAIN 104 148 OMPA-LIKE.
SQ SEQUENCE 177 AA; 18839 MW; A3AB53402ECB1ADD CRC64;

Query Match 19.54; Score 177; DB 1; Length 177;
Best Local Similarity 27.98; Pred. No. 1.7e-07;
Matches 48; Conservative 33; Mismatches 69; Indels 22; Gaps 4;

Oy 2 KRSSFGLVFLVAGCHQDNKTVAGDVSAKTQTAPVTTEPAPEKEBPKEPAPV 61
Db 17 RNPVAVMTLA-LALAGCANKKMPNSAGELGLGAGSA-----TPGS 58

Oy 62 EEKPAVESGTTIASIYDFDKYIKESDQSTDEIVQ-KAKENHMQVLEGTDFGSGSE 120
Db 59 QDDFTVNVGD---RIFFDTSTIRADAQQLQRAQWLRYNYVAITVGHADERTRE 115

Oy 121 YNOALGVKRTLSVKNALVIGKVEKDMIKTISFGETKPKCAQKTRCYKENR 172
Db 116 YNALGARRAAATRDLASQGVPSARMKTSYSGKEKPVAVCDDDISCWSQNR 167

RESULT 10
OM16_BRUME STANDARD; PRT; 168 AA.
AC Q44662;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Outer membrane lipoprotein omp16 precursor (Minor outer membrane
DE protein omp16) (16-kDa OMP) (16.5-kDa minor OMP).
GN OMP16 OR PAL OR BMEI0340 OR BR1695.
OS Brucella melitensis.
OS Brucella suis, and
OS Brucella abortus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459, 29461, 235;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 53-58.
RX SPECIES=B. abortus; STRAIN=544 / Biovar 1;
RX MEDLINE=94341863; PubMed=8063379;
EX Tabor A., Weynants V., Denoel P., Lichfouse B., De Bolle X.,
EX Saman A., Linet J.N., Jettesson J.-J.;
RA "Molecular cloning, nucleotide sequence, and occurrence of a 16.5-
RT kilodalton outer membrane protein of Brucella abortus with similarity
RT to pal lipoproteins."
RL Infect. Immun. 62:3633-3639 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RX SPECIES=B. melitensis; STRAIN=16M / ATCC 23456 / Biotype 1;

```

Query Match 19.4%; Score 176; DB 1; Length 168;
Best Local Similarity 28.1%; Pred. No. 1.9e-07;
Matches 48; Conservative 28; Mismatches 73; Indels 22; Gaps 3;

QY 3 RSVFSEFLVAFLLVAGCSHKMDNKTVDGSAKTQTQTAPVTTEPAPEKEEPQBPVVE 62
DB 9 RPIAIALFMSLAVAGCSAKNLPNNAGDLGLGAGAAATPGSSQDF----- 53
QY 63 EKPAVESGPIIASIYDFDKVKEKSDQETLDEIVQ-KAKENHMQVLEGNDFGSSEY 121
DB 54 ---TWNVGD---RIFEDLSSLRADAQOTLSKQAWLQRYPOYSITTEGHADERTREY 107
QY 122 NQALGVKRTLSVKNALVIKGVKMDIKTISFGETKPKCAQKTRCYCKENRR 172
DB 108 NLALGORRAATRDFLASRGVPTNRRTISYGNRPVAVCDATCWSQNR 158

RESULT 11
OMPA_CITFR
ID_OMPA_CITFR STANDARD; PRT; 238 AA.
AC P24016;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Outer membrane protein A (Outer membrane protein II) (Fragment).
GN OMPA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OS60;
RX MEDLINE=92065252; PubMed=1955870;
RA Lawrence J.G., Ochman H., Hartl D.L.;
RT "Molecular and evolutionary relationships among enteric bacteria.";
RL J. Gen. Microbiol. 137:1911-1921(1991).
CC -! FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
CC STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A
CC RECEPTOR FOR A NUMBER OF T-EVEN LIKE PHAGES. ALSO ACTS AS A PORIN
CC WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
CC SOLUTES (BY SIMILARITY).
CC -! SUBUNIT: Monomer (Probable).
CC -! SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -! SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M63354; AAA23095.1; -
CC PIR; I40703; I40703.
CC HSSP; P02934; IQJP.
CC InterPro; IPR006664; Bac OmpA.
CC InterPro; IPR001035; MotY.
CC InterPro; IPR006665; OmpA/MotB.
CC InterPro; IPR006690; OMPA LIKE.
CC InterPro; IPR000498; OmpA_tmam.
CC Pfam; PF00591; OmpA; 1.
CC Pfam; PF01389; OmpA_membrane; 1.
CC PRINTS; PR01023; NAFLGMOTY.
CC PRINTS; PR01021; OMPADOMAIN.
CC ProDom; PD000930; OmpA/MotB; 1.
CC PROSITE; PS01068; OMPA; 1.
CC Conjugation; Phage recognition; Porin; Outer membrane; Transmembrane;
CC Repeat.
CC NON TER. 1 1
CC TRANSMEM <1 8 POTENTIAL.
CC TRANSMEM 12 27 POTENTIAL.

Tibor A., Aidant N., Letesson J.-J.;
"The tol-pal region in Brucella encodes homologs of the Tol-Pal system
of E. coli.";
Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
RC SPECIES=B.melitensis; STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756688;
RA Delvecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyrides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
Brucella melitensis.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).
RN [4]
SEQUENCE FROM N.A.
RC SPECIES=B.suis; STRAIN=1330 / Biovar 1;
RX MEDLINE=2247741; PubMed=12271122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolony J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
animal and plant pathogens and symbionts.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153 (2002).
RN [5]
CHARACTERIZATION.
RC MEDLINE=99386905; PubMed=10456959;
RX Tibor A., Decelle B., Letesson J.-J.;
RT "Outer membrane proteins Omp10, Omp16, and Omp19 of Brucella spp. are
RT lipoproteins.";
RT Infect. Immun. 67:4960-4962 (1999).
CC -! SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.
CC -! PTM: The N-terminus is blocked.
CC -! SIMILARITY: BELONGS TO THE OMPA FAMILY. PAL SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; I27936; AAA59360.1; -
CC EMBL; AF358662; AAK48919.1; -
CC EMBL; AE009476; AAL51521.1; -
CC EMBL; AF014462; AAN30595.1; -
CC PIR; AF3294; AF3294.
CC PIR; I40346; I40346.
CC TIGR; BR1695; -
CC InterPro; IPR006664; Bac OmpA.
CC InterPro; IPR001035; MotY.
CC InterPro; IPR006665; OmpA/MotB.
CC InterPro; IPR006690; OMPA LIKE.
CC Pfam; PF00591; OmpA; 1.
CC PRINTS; PR01023; NAFLGMOTY.
CC PRINTS; PR01021; OMPADOMAIN.
CC ProDom; PD000930; OmpA/MotB; 1.
CC PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
CC PROSITE; PS01068; OMPA; FALSE NEG.
CC Membrane; Outer membrane; Lipoprotein; signal; Complete proteome.
CC SIGNAL 1 24 PROBABLE.
CC CHAIN 25 168 OUTER MEMBRANE LIPOPROTEIN OMP16.
CC FT LIPID 25 25 N-ACYL DIGLYCERIDE (PROBABLE).
CC DOMAIN 95 139 OMPA-LIKE.
CC SEQUENCE 168 AA; 18232 MW; 17F75F4F05EB1DD8 CRC64;

```

FT TRANSMEM 43 54 POTENTIAL.
FT TRANSMEM 60 76 POTENTIAL.
FT TRANSMEM 82 93 POTENTIAL.
FT DOMAIN 104 109 3 X 2 AA TANDEM REPEATS OF A-P.
FT REPEAT 104 105 1.
FT REPEAT 106 107 2.
FT REPEAT 108 109 3.
FT DOMAIN 158 202 OMPA-LIKE.
FT DISULFID 212 224 BY SIMILARITY.
FT NON TER 238 238
SQ SEQUENCE 238 AA; 25664 MW; 66EA136D73A3F286 CRC64;

Query Match 19.3%; Score 174.5; DB 1; Length 238;
Best Local Similarity 28.7%; Pred. No. 3.9e-07;
Matches 51; Conservative 27; Mismatches 61; Indels 39; Gaps 6;

QY 14 LLVAGCSHKNDKTVAGDVSAKTQVAPVTTPEAPEKEEPKQBPAPVVEKPAVESGTII 73
DB 83 LLSGVGSYRFGQOE-----EAPVVVAPA-----PAEVQTK---HFTLK 119
QY 74 ASIYFDKVEIKESDQETLDEIVQK-----AKENHMQVLLEGNTEFGSSSEYNOALGVK 128
DB 120 SDVLEFNKATLPEGQQALDQYMSQSLNLDPKDGSVVVL--GFTDRIGSDAYNOGLSEK 177
QY 129 RTLSVKVALVKGVEKDMIKTISFGETKP-----KCAQKTRCYKENRRVDVKL 177
DB 178 RAQSVVDYLISKIPSKISARGMGESNPVTGNTCDNVKARAALIDCLAPDRRVEIEV 235

RESULT 12
OMPA_SERMA STANDARD; PRT; 359 AA.
AC P04845;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Outer membrane protein A precursor.
GN OMPA.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85033845; PubMed=6092858;
RA Braun G., Cole S.T.;
RT "DNA sequence analysis of the Serratia marcescens ompA gene: implications for the organisation of an enterobacterial outer membrane protein.";
RL Mol. Gen. Genet. 195;321-328(1984).
CC -!- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A RECEPTOR FOR A NUMBER OF T-EVEN LIKE PHAGES. ALSO ACTS AS A PORIN WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL SOLUTES (BY SIMILARITY).
CC -!- SUBUNIT: Monomer (probable).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EVEL; X00618; CAA25254.1; -.
DR PIR; S07298; S07298.
DR HSSP; P02934; 1QUP.
DR InterPro; IPR005664; Bac OmpA.
DR InterPro; IPR005665; OmpA/MotB.
DR InterPro; IPR005690; OMPA_LIKE.

```

```

DR InterPro; IPR000498; OmpA_tmem.
DR Pfam; PF00691; OmpA; 1.
DR Pfam; PF01389; OmpA membrane; 1.
DR PRINTS; PR01021; OMPADOMAIN.
DR ProDom; PD000930; OmpA/MotB; 1.
DR PROSITE; PS01068; OMPA; 1.
KW Conjugation; Phage recognition; Porin; Outer membrane; Transmembrane;
Signal; Repeat.
FT SIGNAL 1 21
FT CHAIN 22 359 OUTER MEMBRANE PROTEIN A.
FT TRANSMEM 27 40 POTENTIAL.
FT TRANSMEM 62 74 POTENTIAL.
FT TRANSMEM 77 92 POTENTIAL.
FT TRANSMEM 104 114 POTENTIAL.
FT TRANSMEM 118 133 POTENTIAL.
FT TRANSMEM 154 165 POTENTIAL.
FT TRANSMEM 171 187 POTENTIAL.
FT TRANSMEM 193 204 POTENTIAL.
FT DOMAIN 210 219 5 X 2 AA TANDEM REPEATS OF A-P.
FT REPEAT 210 211 1.
FT REPEAT 212 213 2.
FT REPEAT 214 215 3.
FT REPEAT 216 217 4.
FT REPEAT 218 219 5.
FT DOMAIN 268 312 OMPA-LIKE.
FT DISULFID 322 336 BY SIMILARITY.
SQ SEQUENCE 359 AA; 38426 MW; 13992A037C197588 CRC64;

Query Match 19.3%; Score 174.5; DB 1; Length 359;
Best Local Similarity 31.8%; Pred. No. 6.5e-07;
Matches 54; Conservative 25; Mismatches 62; Indels 29; Gaps 6;

QY 24 DNKTVAGDVSAKTQVAPVTTPEAPEKEEPKQBPAPVVEKPAVESGTIIASIFDFDKY 83
DB 191 DNTWLSLGVSYRFGQDDVAPAPAP-----APAPVVEIK-----RFLKSDVLFNFKS 239
QY 84 EIKESDQETLDEIVQK-----AKENHMQVLLEGNTEFGSSSEYNOALGVKRTLSVKVALV 138
DB 240 TLKAEQQALDQLYTLQSSMDPKDGSVVVL--GYTDAVGSQVYNQKLSEQRASQSVVDYLV 297
QY 139 IKGVEKDMIKTISFGE---TKPKCAQKTR-----CYKENRRVDVKL 177
DB 298 SKGIPSDKISARGMGHEADAVTGTCTGYKGRATKAQIVCLAPDRRVEIEV 347

RESULT 13
PAL_ECOLI STANDARD; PRT; 173 AA.
ID PAL_ECOLI
AC P07176;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Peptidoglycan-associated lipoprotein precursor.
GN PAL OR EXCC OR B0741 OR Z0909 OR ECS0776 OR SF0556.
OS Escherichia coli.
OS Escherichia coli. O157:H7, and Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334, 623;
RN [1]
RP SEQUENCE FROM N.A.
RX SPECIES=E.coli; STRAIN=K12; MEDLINE=87133578; PubMed=3545827;
RA Chen R., Henning U.;
RT "Nucleotide sequence of the gene for the peptidoglycan-associated lipoprotein of Escherichia coli K12.";
RL Eur. J. Biochem. 163;73-77(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX SPECIES=E.coli; STRAIN=K12; MEDLINE=92244043; PubMed=1574003;
RA Lazzaroni J.-C., Portallier R.;

```

"The *exsC* gene of *Escherichia coli* K-12 required for cell envelope integrity encodes the peptidoglycan-associated lipoprotein (PAL)".
 Mol. Microbiol. 6:735-742(1992).
 [3]
 SEQUENCE FROM N.A.
 SPECIES=E.coli; STRAIN=X12 / MG1655;
 MEDLINE=9742617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12";
 RL Science 277:1453-1474(1997).
 RN [4]
 SEQUENCE FROM N.A.
 SPECIES=E.coli; STRAIN=X12;
 RC SPECIES=E.coli; STRAIN=X12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map";
 RL DNA Res. 3:137-155(1996).
 RN [5]
 SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=0157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grothbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genomic sequence of enterohaemorrhagic *Escherichia coli* O157:H7";
 RL Nature 409:529-533(2001).
 RN [6]
 SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=0157:H7 / RMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 O157:H7 and genomic comparison with a laboratory strain K-12";
 RL DNA Res. 8:11-22(2001).
 RN [7]
 SEQUENCE OF 1-30 FROM N.A.
 RC SPECIES=E.coli; STRAIN=JM105;
 RX MEDLINE=90078104; PubMed=2687247;
 RA Levensgood S.K., Webster R.B.;
 RT "Nucleotide sequences of the *tolA* and *tolB* genes and localization of
 their products, components of a multistep translocation system in
Escherichia coli";
 RL J. Bacteriol. 171:6600-6609(1989).
 RN [8]
 SEQUENCE FROM N.A.
 RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
 RX MEDLINE=2272406; PubMed=12384590;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.;
 RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity
 through comparison with genomes of *Escherichia coli* K12 and O157";
 RL Nucleic Acids Res. 30:4432-4441(2002).
 RN [9]
 CRYSTALLIZATION.
 RC SPECIES=E.coli;

RX MEDLINE=21111468; PubMed=11173492;
 RA Abergel C., Walburger A., Chenivresse S., Lazdunski C.;
 RT "Crystallization and preliminary crystallographic study of the
 peptidoglycan-associated lipoprotein from *Escherichia coli*";
 RL Acta Crystallogr. D 57:317-319(2001).
 CC -!- FUNCTION: Thought to play a role in bacterial envelope integrity.
 CC Very strongly associated with the peptidoglycan.
 CC -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 anchor.
 CC -!- SIMILARITY: BELONGS TO THE OMPA FAMILY. PAL SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 DR EMBL; X05123; CAA28771.1; -
 DR EMBL; X65796; CAA46673.1; -
 DR EMBL; AE000177; AAC73835.1; -
 DR EMBL; D90713; BAA35407.1; -
 DR EMBL; AE005252; AAG55077.1; -
 DR EMBL; AP02553; BAB34199.1; -
 DR EMBL; M28232; -; NOT ANNOTATED CDS.
 DR EMBL; AE015086; AAN42200.1; ALT_INIT.
 DR PIR; A27534; LPECPG.
 DR PIR; A85577; A85577.
 DR PIR; H90725; H90725.
 DR Ecogene; EG10684; pal.
 DR InterPro; IPR006664; Bac OmpA.
 DR InterPro; IPR006665; OmpA/MotB.
 DR InterPro; IPR006690; OMPA-LIKE.
 DR Pfam; PF00691; OMPA; 1.
 DR PRINTS; PS01021; OMPADOMAIN.
 DR PRODOM; PD000930; OmpA/MotB; 1.
 DR PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
 DR PROSITE; PS01068; OMPA; 1.
 KW Outer membrane; Signal; Lipoprotein; Complete proteome.
 FT CHAIN 1 21 PEPTIDOGLYCAN-ASSOCIATED LIPOPROTEIN.
 FT LIPID 22 173 N-ACYL DIGLYCERIDE.
 FT DOMAIN 105 149 OMPA-LIKE.
 SQ SEQUENCE 173 AA; 18824 MW; 449F9959C0274430 CRC64;
 Query Match 19.0%; Score 172.5; DB 1; Length 173;
 Best Local Similarity 38.8%; Pred. No. 3; e-07;
 Matches 38; Conservative 16; Mismatches 43; Indels 1; Gaps 1;
 Qy 76 IYDFDKYIKESDQETLDEIVQAKEN-HMQVLLGNTDFGSSEYNQALGVKRTLSVK 134
 Db 71 VFYDLKDYIRSDFAQMLDAHANFLRSNPSYKVTVEGHADERTPEYNISLGERRANAVK 130
 Qy 135 NALVIVKVEKMDTKTSFGETKPKCAKTECYKENER 172
 Db 131 MYLQKGVSADQISIVSGKKEPAVLGHDEAAYSKNRR 168
 RESULT 14
 OMPA_SALTY STANDARD; PRT; 350 AA.
 AC P02936;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Outer membrane protein A precursor (Outer membrane protein 33K) (Outer
 DE membrane major heat-modifiable protein).
 GN OMPA OR STM1070.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;

RN SEQUENCE FROM N.A.
 RP MEDLINE=83287368; PubMed=6349993;
 RA Freud R., Cole S.T.;
 RT "Cloning and molecular characterization of the ompA gene from
 RT *Salmonella typhimurium*,"
 RL Eur. J. Biochem. 134:497-502(1983).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Lf2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.B., Spieth J., Clifton S.W., Lettrille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 RT Lf2,"
 RL Nature 413:852-856(2001).
 CC -!- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
 CC STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A
 CC RECEPTOR FOR A NUMBER OF T-EVEN LIKE PHAGES. ALSO ACTS AS A PORIN
 CC WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
 CC SOLUTES (BY SIMILARITY).
 CC SUBUNIT: Monomer (Probable).
 CC -!- SUBUNIT: Monomer (Probable).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -!- SIMILARITY: BELONGS TO THE OMPA FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC ENBL; X02006; CAA26037.1; --
 CC EMBL; A008746; AAL20003.1; --
 CC PIR; A03436; MMEBAT.
 CC HSP; P02934; IQJP.
 CC StyGene; SG10263; ompA.
 CC InterPro; IPR006664; Bac OmpA.
 CC InterPro; IPR001035; MotY.
 CC InterPro; IPR006665; OmpA/MotB.
 CC InterPro; IPR006666; OmpA LIKE.
 CC InterPro; IPR000498; OmpA_tmem.
 CC Pfam; PF00691; OmpA; 1.
 CC Pfam; PF01389; OmpA membrane; 1.
 CC PRINTS; PR01023; NAEFLGMOTY.
 CC PRINTS; PR01021; OMPADOMAIN.
 CC ProDom; PD000930; OmpA/MotB; 1.
 CC PROSITE; PS01068; OmpA; 1.
 CC Conjugation; Phage recognition; Porin; Outer membrane; Transmembrane;
 CC Signal; Repeat; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 350 OUTER MEMBRANE PROTEIN A.
 FT TRANSMEM 27 40
 FT TRANSMEM 55 67 POTENTIAL.
 FT TRANSMEM 70 85 POTENTIAL.
 FT TRANSMEM 97 107 POTENTIAL.
 FT TRANSMEM 111 126 POTENTIAL.
 FT TRANSMEM 146 157 POTENTIAL.
 FT TRANSMEM 163 179 POTENTIAL.
 FT TRANSMEM 185 196 POTENTIAL.
 FT DOMAIN 205 212 4 X 2 AA TANDEM REPEATS OF A-P.
 FT REPEAT 205 206 1.
 FT REPEAT 207 208 2.
 FT REPEAT 209 210 3.
 FT REPEAT 211 212 4.
 FT DOMAIN 261 305 OMPA-LIKE.
 FT DISULFID 315 327 BY SIMILARITY.
 FT CONFLICT 114 114 V -> F (IN REF. 1).
 FT CONFLICT 247 247 S -> I (IN REF. 1).

SQ SEQUENCE 350 AA; 37515 MW; B4AC52C8CDF54FE CRC64;
 Query Match 18.9%; Score 171.5; DB 1; Length 350;
 Best Local Similarity 29.8%; Pred. No. 1.1e-06;
 Matches 53; Conservative 27; Mismatches 59; Indels 39; Gaps 8;
 QY 14 LLVAGCSHKMDNKTIVAGDVSAKTVQTPVTEPAPEKEEPKQEPAPVWEKPAVESGTLI 73
 DB 186 LLSVGVSYRFGQQ-----EAPVVA-PAP-----APAEVQTK-----HFTLK 222
 QY 74 ASIYFDDKYEIKESDOETLDEIVQK-----AKENHNVILLEGTDFGSSSEYNOALGVK 128
 DB 223 SDVLNFENKTLRPEGOALDQLYSQSLNLDPKDGSVVL--GFTDRIGSDATNOGLSEK 280
 QY 129 RTLSVKNALVIKGVKMDKMITISFGTKPKCAQ-----KTR-----ECYKENRVDVVKL 177
 DB 281 RAQSVVDYLISKIPSKISARGMGESNPVTGNTCDNVKFRALIDCLAPDRREVEIEV 338
 RESULT 15
 ID OMPA_ESCHE STANDARD; PRT; 243 AA.
 AC P24754;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Outer membrane protein A (Outer membrane protein II) (Fragment).
 GN OMPA.
 OS Escherichia hermannii.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_taxID=565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33650; and ATCC 33652;
 RX MEDLINE=92065252; PubMed=1955870;
 RA Lawrence J.G., Ochman H., Hartl D.L.;
 RT "Molecular and evolutionary relationships among enteric bacteria,"
 RL J. Gen. Microbiol. 137:1911-1921(1991).
 CC -!- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
 CC STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A
 CC RECEPTOR FOR A NUMBER OF T-EVEN LIKE PHAGES. ALSO ACTS AS A PORIN
 CC WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
 CC SOLUTES (BY SIMILARITY).
 CC SUBUNIT: Monomer (Probable).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -!- SIMILARITY: BELONGS TO THE OMPA FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M63346; AAA24234.1; --
 CC EMBL; M63347; AAA24238.1; --
 CC PIR; IG2386; I62386.
 CC HSP; P02934; 1BXW.
 CC InterPro; IPR006664; Bac OmpA.
 CC InterPro; IPR001035; MotY.
 CC InterPro; IPR006665; OmpA/MotB.
 CC InterPro; IPR006690; OmpA LIKE.
 CC InterPro; IPR000498; OmpA_tmem.
 CC Pfam; PF00691; OmpA; 1.
 CC Pfam; PF01389; OmpA membrane; 1.
 CC PRINTS; PR01023; NAEFLGMOTY.
 CC PRINTS; PR01021; OMPADOMAIN.
 CC ProDom; PD000930; OmpA/MotB; 1.
 CC PROSITE; PS01068; OmpA; 1.
 CC Conjugation; Phage recognition; Porin; Outer membrane; Transmembrane;
 CC Repeat.

```

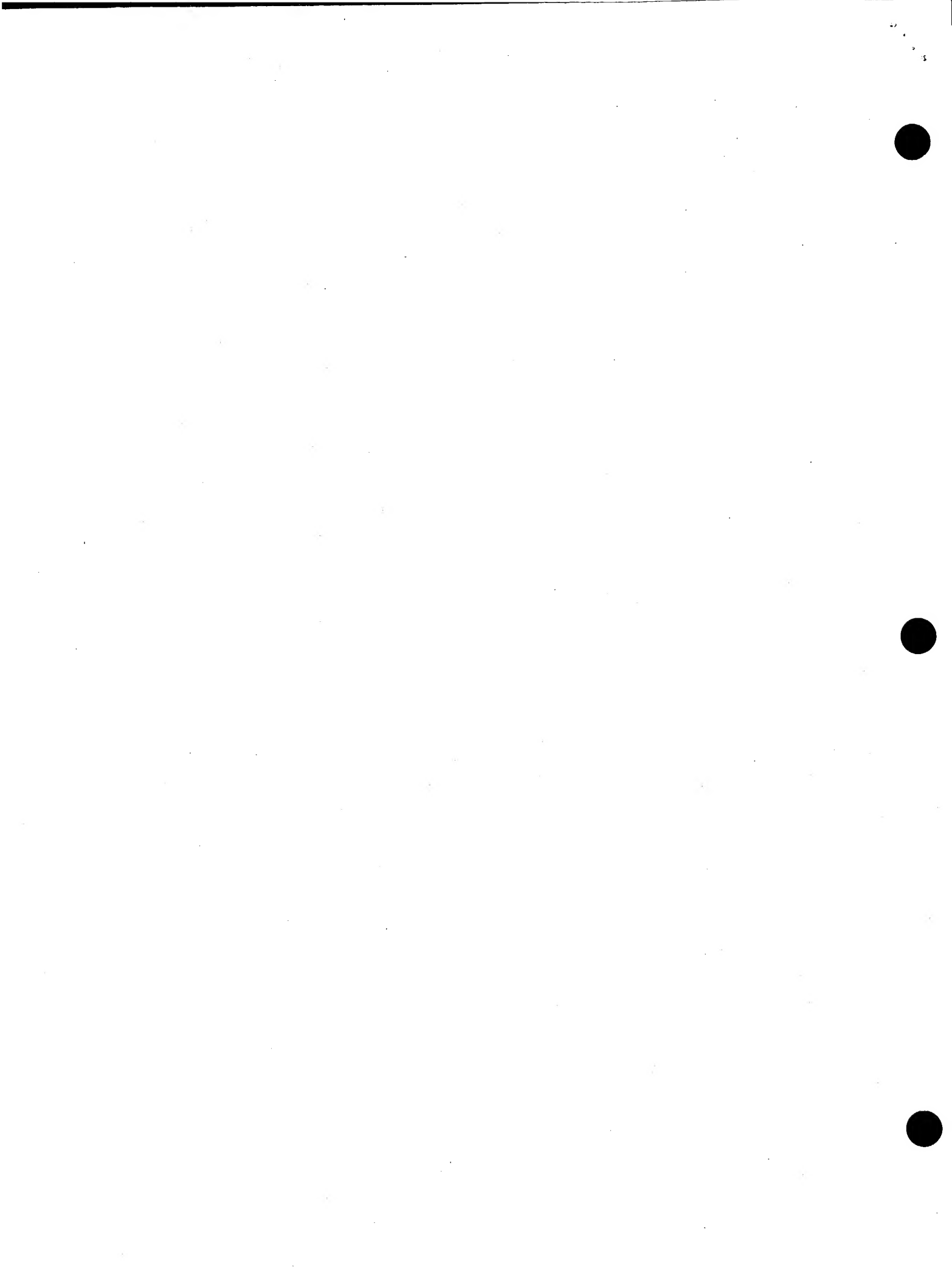
FT NON TER 1 1
FT TRANSMEM <1 8
FT TRANSMEM 12 27
FT TRANSMEM 47 58
FT TRANSMEM 64 80
FT TRANSMEM 86 97
FT TRANSMEM 107 114
FT TRANSMEM 107 108
FT TRANSMEM 109 110
FT TRANSMEM 111 112
FT TRANSMEM 113 114
FT TRANSMEM 163 207
FT TRANSMEM 217 229
FT TRANSMEM 243 243
SQ SEQUENCE 243 AA; 26202 MW; AA6CB6BA390D1E9D CRC64;

Query Match 18.9%; Score 171; DB 1; Length 243;
Best Local Similarity 30.3%; Pred No 7.7e-07;
Matches 54; Conservative 26; Mismatches 60; Indels 38; Gaps 8;

Qy 14 LLVAGCSHKMDNKTAVGDSAKTVQTAPVTEPAPEKEBFPAPVVEEKPAVESGTII 73
Db 87 LLSVGSYSRFGQCEAA-----APVVA-EAP-----APAPEVQTK---HFTLK 124
Qy 74 ASIVFDKYEIKESDOETLDEIVQK-----AKENHMQVLLEGTDFGSSSEYNOALGVK 128
Db 125 SDVLFNFNKATLPEGQALDQMYTQLSNLDPKDGSVWL--GFTDRIGSDAYNQGLSEK 182
Qy 129 RTLGVKNALVIKGVKMDIKTISFGETKPKCAQ-----KTR-----ECYKENRRVDVKL 177
Db 183 RAQSVWDYLIKGIKIPSDKISARGMGESNPVTGNTCDNVKPRALIDCLAPDRRVIEV 240

```

Search completed: December 10, 2003, 18:34:32
 Job time : 9.0288 secs



GenCore version 5.1.6
 Copyright (c) 1993 - 2003 Compugen Ltd.
 OM protein - protein search, using sw model
 Run on: December 10, 2003, 18:29:51 ; Search time 26.0065 Seconds
 (without alignments)
 1776.146 Million cell updates/sec
 Title: US-10-080-113-2
 Perfect score: 906
 Sequence: 1 MKRSSVGFSLVAFLLVAGS.....AQKTRECYKENRVDVLMK 179

Scoring table: BLOSUM62

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : SPTREMBL_23.*
 1: sp_archaea.*
 2: sp_bacteria.*
 3: sp_fungi.*
 4: sp_human.*
 5: sp_invertebrate.*
 6: sp_mammal.*
 7: sp_mhc.*
 8: sp_organelle.*
 9: sp_phage.*
 10: sp_plant.*
 11: sp_rodent.*
 12: sp_virus.*
 13: sp_vertibrate.*
 14: sp_unclassified.*
 15: sp_rvirus.*
 16: sp_bacteriap.*
 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	906	100.0	179	16 Q25750	Q25750 helicobacte
2	877	96.8	179	16 Q92K87	Q92K87 helicobacte
3	836	92.3	179	2 Q9ZAW7	Q9ZAW7 helicobacte
4	793	87.5	157	2 Q9RNF4	Q9RNF4 helicobacte
5	346	38.2	68	2 Q8KRH1	Q8KRH1 helicobacte
6	269	29.7	153	2 Q9X691	Q9X691 helicobacte
7	257.5	28.4	165	2 Q46099	Q46099 campylobact
8	250	27.6	165	16 Q46123	Q46123 campylobact
9	222.5	24.6	186	16 Q9PC85	Q9PC85 xylella fas
10	219	24.2	172	16 Q8Y1F4	Q8Y1F4 ralstonia s
11	215.5	23.8	326	2 O06895	O06895 campylobact
12	209.5	23.1	319	16 Q9PMI4	Q9PMI4 campylobact
13	203.5	22.5	323	2 Q9XCZ5	Q9XCZ5 pseudomonas
14	201.5	22.2	218	16 Q8Y0Z3	Q8Y0Z3 ralstonia s
15	200	22.1	170	16 Q8D2E3	Q8D2E3 wigglewort
16	198	21.9	188	16 Q9A3H5	Q9A3H5 caulobacter

17	195.5	21.6	321	2	Q9XCZ6	Q9XCZ6 pseudomonas
18	195.5	21.6	328	2	Q9XCZ7	Q9XCZ7 pseudomonas
19	194.5	21.5	321	2	Q9XCZ3	Q9XCZ3 pseudomonas
20	193	21.3	168	16	Q914Z4	Q914Z4 pseudomonas
21	192	21.2	168	2	Q51489	Q51489 pseudomonas
22	191	21.1	210	16	Q914T3	Q914T3 pseudomonas
23	189.5	20.9	181	16	Q8DAM2	Q8DAM2 vibrio vuln
24	185	20.4	170	16	Q8KEP8	Q8KEP8 chlorobium
25	183.5	20.3	318	2	Q9X4S1	Q9X4S1 pseudomonas
26	183	20.2	157	2	Q47958	Q47958 haemophilus
27	180.5	19.9	130	2	Q9L3Q8	Q9L3Q8 pasteurilla
28	180.5	19.9	287	2	Q9X4B1	Q9X4B1 campylobact
29	180.5	19.9	315	16	Q8SEH8	Q8SEH8 fusobacteri
30	180	19.9	172	16	Q8PEF3	Q8PEF3 xanthomonas
31	179.5	19.8	287	2	Q9X4B0	Q9X4B0 campylobact
32	178	19.6	152	2	Q44157	Q44157 actinobacil
33	178	19.6	168	16	Q8ZGZ0	Q8ZGZ0 versinia pe
34	178	19.6	181	16	Q8PHV8	Q8PHV8 xanthomonas
35	177	19.5	155	16	Q9ZCH2	Q9ZCH2 rickettsia
36	176	19.4	202	16	Q9PJE0	Q9PJE0 chlamydia m
37	174.5	19.3	177	16	Q8EDJ9	Q8EDJ9 shewanella
38	172.5	19.0	367	2	Q9RM69	Q9RM69 erwinia car
39	171.5	18.9	151	16	Q8R6J1	Q8R6J1 fusobacteri
40	171.5	18.9	157	16	Q8RGN8	Q8RGN8 fusobacteri
41	171.5	18.9	181	16	Q8CW95	Q8CW95 escherichia
42	171.5	18.9	350	16	Q8Z7S0	Q8Z7S0 salmonella
43	170.5	18.8	172	16	Q8SGG3	Q8SGG3 fusobacteri
44	170	18.8	174	16	Q8XPF6	Q8XPF6 salmonella
45	170	18.8	325	2	Q9X4R8	Q9X4R8 pseudomonas

ALIGNMENTS

RESULT 1
 Q25750
 ID Q25750 PRELIMINARY; PRT; 179 AA.
 AC Q25750;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Peptidoglycan associated lipoprotein precursor (OMP18).
 GN HP1125.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185;
 RA Tomb J.-F., White O., Kexlavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 pylori.";
 RL Nature 388:539-547(1997).
 DR EMBL; AE000619; AAD08169.1; -.
 DR TIGR; HP1125; -.
 DR InterPro; IPR006664; Bac OmpA.
 DR InterPro; IPR006665; OmpA/MotB.
 DR Pfam; PF00691; OmpA; 1.
 DR PRINTS; PRO1021; OMPADOMAIN.
 DR ProDom; PD000930; OmpA/MotB; 1.
 KW Hypothetical protein; Lipoprotein; Complete proteome.
 SQ SEQUENCE 179 AA; 19978 MW; 5B5521E284E27B4C CRC64;

```
Query Match 100.0%; Score 906; DB 16; Length 179;
Best Local Similarity 100.0%; Pred. No. 1.8e-64;
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKRSSVFLVAFLLVAGCSHKMDNKTIVAGDVSAKTIVQTAPVTEPAPEKEPKQEPAPV 60
    |||||
Db 1 MKRSSVFLVAFLLVAGCSHKMDNKTIVAGDVSAKTIVQTAPVTEPAPEKEPKQEPAPV 60
    |||||

Qy 61 VEKPAVESGTTIASIYFDKYEIKESDQETLDEIVQKAKENHMOVLLEGNTEFGSSE 120
    |||||
Db 61 VEKPAVESGTTIASIYFDKYEIKESDQETLDEIVQKAKENHMOVLLEGNTEFGSSE 120
    |||||

Qy 121 YNQALGVKRTLSVKNALVIKGVKMDKTIISFGETPKCAQKTRCYKRNRRVDVKLMK 179
    |||||
Db 121 YNQALGVKRTLSVKNALVIKGVKMDKTIISFGETPKCAQKTRCYKRNRRVDVKLMK 179
    |||||

RESULT 2
Q9ZK97 ID Q9ZK97 PRELIMINARY; PRT; 179 AA.
AC Q9ZK97; 1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE Putative outer membrane protein.
GN JHP1054.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uric-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovlis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
DR EMBL; AE001533; AAD06833.1; -.
DR InterPro; IPR006664; Bac_OmpA.
DR Pfam; PF00691; OmpA; 1.
DR PRINTS; PR01021; OMPADOMAIN.
DR ProDom; PD000930; OmpA/MotB; 1.
KW Complete proteome.
SQ SEQUENCE 179 AA; 20030 MW; F62BB6191864598D CRC64;

Query Match 96.8%; Score 877; DB 16; Length 179;
Best Local Similarity 95.5%; Pred. No. 3.7e-62;
Matches 171; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MKRSSVFLVAFLLVAGCSHKMDNKTIVAGDVSAKTIVQTAPVTEPAPEKEPKQEPAPV 60
    |||||
Db 1 MKRSSVFLVAFLLVAGCSHKMDNKTIVAGDVSAKTIVQTAPVTEPAPEKEPKQEPAPV 60
    |||||

Qy 61 VEKPAVESGTTIASIYFDKYEIKESDQETLDEIVQKAKENHMOVLLEGNTEFGSSE 120
    |||||
Db 61 VEKPAVESGTTIASIYFDKYEIKESDQETLDEIVQKAKENHMOVLLEGNTEFGSSE 120
    |||||

Qy 121 YNQALGVKRTLSVKNALVIKGVKMDKTIISFGETPKCAQKTRCYKRNRRVDVKLMK 179
    |||||
Db 121 YNQALGVKRTLSVKNALVIKGVKMDKTIISFGETPKCAQKTRCYKRNRRVDVKLMK 179
    |||||

RESULT 3
Q9ZAW7 ID Q9ZAW7 PRELIMINARY; PRT; 179 AA.
AC Q9ZAW7; 1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
```

```
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE OMP22.
GN OMP22.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
STRAIN=KCTC0217BP;
RA Seo W.Y., Kim J.S., Jang J.H., Yu G.J., Yum J.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U75869; AAD09577.1; -.
DR InterPro; IPR006664; Bac_OmpA.
DR InterPro; IPR006665; OmpA/MotB.
DR Pfam; PF00691; OmpA; 1.
DR PRINTS; PR01021; OMPADOMAIN.
DR ProDom; PD000930; OmpA/MotB; 1.
SQ SEQUENCE 179 AA; 20024 MW; 025ADC298F50C5AC CRC64;

Query Match 92.3%; Score 836; DB 2; Length 179;
Best Local Similarity 91.6%; Pred. No. 6.6e-59;
Matches 164; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MKRSSVFLVAFLLVAGCSHKMDNKTIVAGDVSAKTIVQTAPVTEPAPEKEPKQEPAPV 60
    |||||
Db 1 MKRSSVFLVAFLLVAGCSHKMDNKTIVAGDVSAKTIVQTAPVTEPAPEKEPKQEPAPV 60
    |||||

Qy 61 VEKPAVESGTTIASIYFDKYEIKESDQETLDEIVQKAKENHMOVLLEGNTEFGSSE 120
    |||||
Db 61 VEKPAVESGTTIASIYFDKYEIKESDQETLDEIVQKAKENHMOVLLEGNTEFGSSE 120
    |||||

Qy 121 YNQALGVKRTLSVKNALVIKGVKMDKTIISFGETPKCAQKTRCYKRNRRVDVKLMK 179
    |||||
Db 121 YNQALGVKRTLSVKNALVIKGVKMDKTIISFGETPKCAQKTRCYKRNRRVDVKLMK 179
    |||||

RESULT 4
Q9RNF4 ID Q9RNF4 PRELIMINARY; PRT; 157 AA.
AC Q9RNF4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE PAL.
GN EXCC.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
STRAIN=RU-1;
RA Kolesnikov T., Gekas S., Lee A.;
RT "Identification of Helicobacter pylori Antigens.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF191881; AAF04276.1; -.
DR InterPro; IPR006664; Bac_OmpA.
DR InterPro; IPR006665; OmpA/MotB.
DR Pfam; PF00691; OmpA; 1.
DR PRINTS; PR01021; OMPADOMAIN.
DR ProDom; PD000930; OmpA/MotB; 1.
SQ SEQUENCE 157 AA; 17536 MW; E77FB05D3C4E5641C CRC64;

Query Match 87.5%; Score 793; DB 2; Length 157;
Best Local Similarity 99.4%; Pred. No. 1.5e-55;
Matches 156; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 23 MDNKTIVAGDVSAKTIVQTAPVTEPAPEKEPKQEPAPVVEKPAVESGTTIASIYFDK 82
    |||||
Db 1 MDNKTIVAGDVSAKTIVQTAPVTEPAPEKEPKQEPAPVVEKPAVESGTTIASIYFDK 60
    |||||

Qy 83 YEIKESDQETLDEIVQKAKENHMOVLLEGNTEFGSSEYNQALGVKRTLSVKNALVIKGV 142
    |||||
```

Db 61 YEIKESDQETLDELVRKAKENHMQVLLGNTDFGSSSEYNQALGVKRTLSVKNALVIKGV 120
QY 143 EKDWIKTISFGETPKCAQKTRCYKRNRRVDVKLMK 179
Db 121 EKDWIKTISFGETPKCAQKTRCYKRNRRVDVKLMK 157
RESULT 5
Q9X691 PRELIMINARY; PRT; 68 AA.
AC Q9X691
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Hpl125 (Fragment)
GN Hpl125
OS Helicobacter pylori.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VNH-85;
RA Hanh N.T.H., Son L.B., Hoa T.Q., Anh H.T.;
RT "Gene hpl125 of Helicobacter pylori, clinical isolate from a
RL Vietnamese patient";
DR EMBL; AF534874; AA03483.1; -;
FT NON_TER 1
FT NON_TER 68
SQ SEQUENCE 68 AA; 7425 MW; 34ASB283CG35A763 CRC64;

Query Match 38.2%; Score 346; DB 2; Length 68;
Best Local Similarity 100.0%; Pred. No. 1.8e-20;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 23 MDNKTAGDVSAKTQVAPVTTEPAPEKEPKQEPAPVVEKPAVSGTIIASIVYDFDK 82
Db 1 MDNKTAGDVSAKTQVAPVTTEPAPEKEPKQEPAPVVEKPAVSGTIIASIVYDFDK 60
QY 83 YEIKESDQ 90
Db 61 YEIKESDQ 68

RESULT 6
Q9X691 PRELIMINARY; PRT; 153 AA.
AC Q9X691
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Membrane-associated protein map18.
GN MAP18.
OS Helicobacter hepaticus.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=32025;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MU;
RX MEDLINE=99403277; PubMed=10473529;
RA Livingston R.S., Riley L.K., Hook R.R. Jr., Besch-Williford C.L.,
RA Franklin C.L.;
RT "Cloning and expression of an immunogenic membrane-associated protein
RT assay";
RL Clin. Diagn. Lab. Immunol. 6:745-750(1999).
RL EMBL; AF134212; AAD30109.1; -;
DR InterPro; IPR006664; Bac OmpA.
DR InterPro; IPR006665; OmpA/MotB.
DR Pfam; PF00691; OmpA; 1.

DR PRINTS; PR01021; OMPADOMAIN.
DR ProDom; PD000930; OmpA/MotB; 1.
SQ SEQUENCE 153 AA; 16118 MW; AF43D36BD18C6EEE CRC64;
Query Match 29.7%; Score 269; DB 2; Length 153;
Best Local Similarity 48.1%; Pred. No. 6.2e-14;
Matches 50; Conservative 19; Mismatches 35; Indels 0; Gaps 0;
QY 76 IYDFDKYKESDQETLDEIVQKAKENHMQVLLGNTDFGSSSEYNQALGVKRTLSVKN 135
Db 50 VLDFDKYDIRSDMEDRVEKSAALKSTGAKVVLGHTSYGSDAYNVALGTKANAVKN 109
QY 136 ALVTKGVKEMIKTISFGETPKCAQKTRCYKRNRRVDVKLMK 179
Db 110 ALTRGVNASQIKTVSYGSKPTCTSDTPECNOENRRRVEFKLAK 153

RESULT 7
Q46099 PRELIMINARY; PRT; 165 AA.
AC Q46099
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Omp18 protein.
GN PAL OR CJAD.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M275;
RX MEDLINE=96201603; PubMed=8613402;
RA Konkel M.E., Mead D.J., Clephak W. Jr.;
RT "Cloning, sequencing, and expression of a gene from Campylobacter
RT jejuni encoding a protein (Omp18) with similarity to peptidoglycan-
RT associated lipoproteins";
RL Infect. Immun. 64:1850-1853(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=72DZ/92;
RA Pawelec D.P., Jagusztyn-Krynicka K.E.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; U47617; AAC35420.1; -;
DR EMBL; AJ132802; CAA10786.1; -;
DR InterPro; IPR006664; Bac OmpA.
DR InterPro; IPR006665; OmpA/MotB.
DR Pfam; PF00691; OmpA; 1.
DR PRINTS; PR01021; OMPADOMAIN.
DR ProDom; PD000930; OmpA/MotB; 1.
KW Plasmid; Signal.
SQ SEQUENCE 165 AA; 17890 MW; 35B67427864BBC9D CRC64;

Query Match 28.4%; Score 257.5; DB 2; Length 165;
Best Local Similarity 33.9%; Pred. No. 5.6e-13;
Matches 60; Conservative 37; Mismatches 61; Indels 19; Gaps 5;
QY 6 VFSELVAF-LLVAGCGSHMDNKTAVGDSAKTVQTPVTEPAPEKEPKQEPAPVVEK 64
Db 5 LFSSIAAFALVTSQCTK--STSVSGDTSDS-----NRGTGSGDGDIDSK 49
QY 65 PAVESGTIIASIVYDFDKYKESDQETL--DRIVQKAKENHMQVLLGNTDFGSSSEYN 122
Db 50 ISQINDT-LGVYDFDKFNIRPDMQNVNNTANINFNNEVSGVSTVEGNCDEWGTDEYN 108
QY 123 QALGVKRTLSVKNALVIKGVKEMIKTISFGETPKCAQKTRCYKRNRRVDVKLMK 179
Db 109 QALGLKRAKAVKEALIAQGVNSDRIAVKSYGNETPNVCTEKTACDAQNRRRBFKLRS 165

RESULT 8

Q46123 PRELIMINARY; PRT; 165 AA.

AC Q46123; 1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE PEPTIDOGLYCANE associated lipoprotein precursor (Peptidoglycan associated lipoprotein) (OMP18).

OS OMP18 OR PAL OR CJO113.

GN Campylobacter jejuni.

OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales; Campylobacteraceae; Campylobacter.

OX NCBI_TaxID=197;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 29428;

RX MEDLINE=96121222; PubMed=8576327;

RA Burnens A., Stucki U., Nicolet J., Frey J.;

RT "Identification and characterization of an immunogenic outer membrane protein of Campylobacter jejuni.";

RL J. Clin. Microbiol. 33:2826-2832 (1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=NCTC 11168;

RX MEDLINE=20150912; PubMed=10688204;

RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C., Basham D., Chillingworth T., Davies R.M., Feitwell T., Holtroyd S., Jagers K., Kariyasev A.V., Moule S., Pallen M.J., Penn C.W., Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M., Whitehead S., Barrell B.G.;

RT "The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences.";

RL Nature 403:665-668 (2000).

DR ENBL; X63374; CAAS8288.1; -

DR InterPro; IPR006664; Bac OmpA.

DR Pfam; PF00691; OmpA; 1.

DR PRINTS; PR01021; OMPADOMAIN.

DR ProDom; PD000930; OmpA/MotB; 1.

FT SIGNAL 1 18 POTENTIAL.

FT CHAIN 19 165 PEPTIDOGLYCANE ASSOCIATED LIPOPROTEIN.

SQ SEQUENCE 165 AA; 17829 MW; 24CA7928532CB66 CRC64;

Query Match 27.6%; Score 250; DB 16; Length 165;

Best Local Similarity 32.6%; Pred. No. 2.2e-12;

Matches 59; Conservative 36; Mismatches 68; Indels 18; Gaps 4;

QY 1 MKRSSVFLVAFLLVAGCSHKMDKNTVAGDSYSAKTVQAPVTPEAPEKEEPKQEPAPV 60

DB 1 MKKILETSIAALAVISGSKTK--STVSGGSVDVS-----NRGSGSGDWD 45

QY 61 VEEKPAVESGTTIASIYDFDKYEIKESDQETL--DEIVQAKENHMQVLLEGNTDFGS 118

DB 46 IDSKISQLNDT-LNKVYDFDKFNIRPDQNVVSTNANIFNTEVSGVSITVEGNCDEWGT 104

QY 119 SEYNQALGKRTLSVKNALVIGKVEKDMKITSGETKPKCAOKTRCYKENARVDVKLM 178

DB 105 DEYNQALGKRAKAVEALIAKGNVADRIAVKSYGETNPFVCTKACDAQNRAEPKLS 164

QY 179 K 179

DB 165 R 165

RESULT 9

Q9PC85 PRELIMINARY; PRT; 186 AA.

AC Q9PC85; (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Outer membrane protein P6 precursor.

GN XF1896.

OS Xylella fastidiosa.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae; Xylella.

OX NCBI_TaxID=2371;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=9a5c;

RX MEDLINE=20365717; PubMed=10910347;

RA Simpson A.J.G., Rehnach F.C., Arruda P., Abreu F.A., Acencio M., Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S., Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S., Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H., Colauto N.B., Colombo C., Costa F.P., Costa M.C.R., Costa-Neto C.M., Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H., Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A., Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R., Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A., Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper B.L., Kitajima J.P., Krieger J.E., Kuramae E.F., Laigret F., Lambais M.R., Leite L.C.C., Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A., Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L., Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y., Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B., Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S., Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A., de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A., Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B., Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M., de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E., da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr., da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A., de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H., Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L., Zago M.A., Zatz M., Meidanis J., Setubal J.C.;

RT "The genome sequence of the plant pathogen Xylella fastidiosa.";

RL Nature 406:151-159 (2000).

DR ENBL; AE004009; AAF84702.1; -

DR InterPro; IPR006664; Bac OmpA.

DR Pfam; PF00691; OmpA; 1.

DR PRINTS; PR01021; OMPADOMAIN.

DR ProDom; PD000930; OmpA/MotB; 1.

KW Complete proteome.

SQ SEQUENCE 186 AA; 20431 MW; C8DD2F6233DB0C92 CRC64;

Query Match 24.6%; Score 222.5; DB 16; Length 186;

Best Local Similarity 32.4%; Pred. No. 3.9e-10;

Matches 61; Conservative 25; Mismatches 67; Indels 35; Gaps 7;

QY 3 RSSVPSFLVAFLLVAGCSHKMDKNTVAGDSYSAKTVQAPVTPEAPEKEEPKQEPAPV 62

DB 14 RFLISLSTLSALVA-CSKKVEQP---HVPVTKM--APTSTPTPTTPTD----- 60

QY 63 EKPAVESGTTIAS-----IYDFDKYEIKESDQETL---EIVQAKENHMQV 107

DB 61 -----SSGLYTAADLDTDAICLRQVVYDFDKDVKKEFTVLGCHAKYLRNPSAH--I 113

QY 108 LLENTDFGSGSEYNQALGKRTLSVKNALVIGKVEKDMKITSGETKPKCAOKTRCY 167

DB 114 TLOGNTDGRSREYNIALGERGRNSVLSLOANGASSQLNVVSGEERPVCTESTESCW 173

QY 168 KENERVDV 175

DB 174 SRNRVEI 181

RESULT 10

Q8Y1F4 PRELIMINARY; PRT; 172 AA.

ID Q8Y1F4

AC Q8Y1F4; (TrEMBLrel. 15, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

```

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Probable peptidoglycan-associated lipoprotein precursor.
GN PAL OR RS00736 OR RS05117.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chaudier M., Choise N., Claudel-Renard C., Cunac S., Denance N.,
RA Gaspin C., Lavi M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siguer P., Thebaud P., Whalen R., Wincker P., Levy M.,
RA Weissbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646060; CAD14266.1; -.
DR InterPro; IPR006664; Bac_OmpA.
DR InterPro; IPR006665; OmpA/MotB.
DR InterPro; IPR006690; OmpA_LIKE.
DR Pfam; PF00691; OmpA; 1.
DR PRINTS; PR01021; OMPADOMAIN.
DR ProDom; PD000930; OmpA/MotB; 1.
DR PROSITE; PS01068; OmpA; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 172 AA; 18656 MW; 5CFF28B6AE7702D9 CRC64;

Query Match 24.2%; Score 219; DB 16; Length 172;
Best Local Similarity 32.4%; Pred. No. 6.8e-10;
Matches 56; Conservative 29; Mismatches 66; Indels 22; Gaps 5;

QY 10 LVAFLLVAGGSH--KMD---NKTVAGDVSAKTVPVTEPAPEKPEKQEPAPVVEE 63
D 13 IAALLGACSGSVKLDLDTNATGGAAGADTENTPVDV-----SRDELTDENSLAKE 68
QY 64 KPAVESGTTIASIYDFDKYKESDQETLDEIVQAK-ENHMOVLLEGNTEBFGSSEYN 122
D 69 -----SVYDFSDTVKPEYQGLTQHARVYQSHNQKRVLIQGNITDERTGSEYN 117
QY 123 QALGVKRTLSVKNALVKGVEKDMKTIISFGETKPKCAQKTRCYKENRRVDV 175
D 118 LALGQKAEAVRALSLSLGVPDQMSVSLGKPKFQASGHDESWAQNRSDI 170

RESULT 11
ID O06895 PRELIMINARY; PRT; 326 AA.
AC O06895;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Cadp precursor (Outer membrane protein).
GN CADP.
OS Campylobacter coli.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=195;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M275;
RX MEDLINE=97363710; PubMed=3220003;
RA Konkel M.E., Garvis S.G., Tipton S.L., Anderson D.E. Jr.,
RA Cieplak W. Jr.;
RT "Identification and molecular cloning of a gene encoding a
fibronectin-binding protein (Cadp) from Campylobacter jejuni.";
RL Mol. Microbiol. 24:953-963(1997).
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE OMP25 / ROPS FAMILY.
DR EMBL; U87559; AAC35418.1; -.

```

```

DR InterPro; IPR006664; Bac_OmpA.
DR InterPro; IPR006665; OmpA/MotB.
DR InterPro; IPR006690; OmpA_LIKE.
DR InterPro; IPR000498; OmpA_tmem.
DR Pfam; PF00691; OmpA; 1.
DR Pfam; PF01389; OmpA membrane; 1.
DR PRINTS; PR01021; OMPADOMAIN.
DR ProDom; PD000930; OmpA/MotB; 1.
DR PROSITE; PS01068; OmpA; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 326 CADP.
SQ SEQUENCE 326 AA; 36744 MW; D38ACC3DD294232 CRC64;

Query Match 23.8%; Score 215.5; DB 2; Length 326;
Best Local Similarity 35.2%; Pred. No. 2.7e-09;
Matches 56; Conservative 23; Mismatches 65; Indels 15; Gaps 4;

QY 26 KTVAGDVSAKTVPVTEPAPEKPEKQEPAPVVEEKPAV--ESG---TIIASIIYDF 80
D 175 KVAEQVKEVAIE-----PRVAVPTCSQCPAEPREGAMLDENGCKTISFEHGF 225
QY 81 DKYEIKESDQETLDEIVQAKEN-HMOVLLEGNTEBFGSSEYNQALGVKRTLSVKNALVI 139
D 226 DKVDINPAFEKIKETIAQLDENARVDTILEGHTDNIGSRAYNQKLSERRAASVAKLEK 285
QY 140 KGVEKDMKTIISFGETKPKCAQKTRCYKENRRVDVVKLM 178
D 286 FGVDKDRIQTVGVGQDKPRSRNETKEGRADNRVDKAEI 324

RESULT 12
ID Q9PMI4 PRELIMINARY; PRT; 319 AA.
AC Q9PMI4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane fibronectin-binding protein.
GN CADF OR CUI478C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshav A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139078; CAB73900.1; -.
DR InterPro; IPR006664; Bac_OmpA.
DR InterPro; IPR001035; MotY.
DR InterPro; IPR006665; OmpA/MotB.
DR InterPro; IPR006690; OmpA_LIKE.
DR InterPro; IPR000498; OmpA_tmem.
DR Pfam; PF00691; OmpA; 1.
DR Pfam; PF01389; OmpA membrane; 1.
DR PRINTS; PR01023; NAFLGMOTY.
DR PRINTS; PR01021; OMPADOMAIN.
DR ProDom; PD000930; OmpA/MotB; 1.
DR PROSITE; PS01068; OmpA; 1.
KW Complete proteome.
SQ SEQUENCE 319 AA; 35997 MW; 464EFE7243DD683 CRC64;

Query Match 23.1%; Score 209.5; DB 16; Length 319;
Best Local Similarity 39.3%; Pred. No. 8e-09;

```

Matches 55; Conservative 19; Mismatches 59; Indels 7; Gaps 4;

QY 45 EPAPEKEPKQAPVVEKPAV--ESG---TIIASIVDFDKYIEKESQDETLEIVQK 99
 Db 179 EVADTRATE-QAKCFVPEREGALLDENGCEKTSISLEGHFGDKTTINPTQEKIKEIAKV 237

QY 100 AKEN-HMQVLLGNTDEFGSSRYNAGLVKRTLSVKNALVIKGVKDKMKTISFGTKPK 158
 Db 238 LBNERYDITLGHNDNGSRAYNOKLSERRAKSVANELEKYGVEKRITVIGQDNPR 297

QY 159 CAQKTRCYKENRRVDVKLM 178
 Db 298 SSNDTKEGRADNRVDKFI 317

RESULT 13
 Q9XCZ5 PRELIMINARY; PRT; 323 AA.

AC Q9XCZ5;
 DT 01-NOV-1999 (TremBLrel. 12, Created)
 DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
 DE Outer membrane protein OprF (Fragment).
 GN OPRF.
 CS Pseudomonas stutzeri CMT.9.A.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=91505;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 4066;
 RX MEDLINE=99319345; PubMed=10390872;
 EX Vermeiren H., Willems A., Schoofs G., de Mot R., Keijers V., Hai W.,
 RA Vanderleyden J.;
 RA "The rice inoculant strain Alcaligenes faecalis A15 is a nitrogen-
 RT fixing Pseudomonas stutzeri.";
 RL Syst. Appl. Microbiol. 22:215-224 (1999).
 DR EMBL; AF117973; RAD39368.1; -
 DR InterPro; IPR006665; OMPA/MotB.
 DR Pfam; PR00691; OMPA; 1.
 DR ProDom; PD000930; OMPA/MotB; 1.
 DR PROSITE; PS01068; OMPA; 1.
 FT NON TER 323 323
 SQ SEQUENCE 323 AA; 34807 MW; 142EF230B4CDA559 CRC64;

Query Match 22.5%; Score 203.5; DB 2; Length 323;
 Best Local Similarity 35.7%; Pred. No. 2.4e-08;
 Matches 51; Conservative 30; Mismatches 47; Indels 15; Gaps 6;

QY 38 QPAPVTEPAPEKEPKQAPV--EKPAPVSGTIIASIVDFDKYIEKE---SDQET 92
 Db 189 QVAPV--EPTPE-----PAPAPVDTPEPAPEVVRVELDKVDFDKSRVRESYSIDKN 241

QY 93 LDEIVQKAKENHMQVLLGNTDEFGSSRYNAGLVKRTLSVKNALVIKGVKDKMKTIS 151
 Db 242 LADFVQVQPT--TTVVGHTDSVGTQINQRLSERRAEAVRNLVNGVQGNRVNSVG 299

QY 152 FGETPKCAQKTRCYKENRRVD 174
 Db 300 YGESRPVADNSTEGRIQNRVE 322

RESULT 14
 Q8Y0Z3 PRELIMINARY; PRT; 218 AA.

AC Q8Y0Z3;
 DT 01-MAR-2002 (TremBLrel. 20, Created)
 DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
 DE Probable signal peptide protein.
 GN RSC0900 OR RS04515.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).

OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Ralstoniaceae; Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguanave F., Gouzy J., Mangenot S.,
 RA Arlat M., Billault A., Bottier P., Camus J.C., Cattolico L.,
 RA Chandler M., Choisme N., Claudel-Renard C., Cumnac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
 RL Nature 415:497-502 (2002).
 DR EMBL; AL646081; CAD14602.1; -
 DR InterPro; IPR006664; Bac_OmpA.
 DR InterPro; IPR006665; OmpA/MotB.
 DR Pfam; PF00691; OmpA; 1.
 DR PRINTS; PR01021; OMPADOMAIN.
 DR ProDom; PD000930; OmpA/MotB; 1.
 KW Complete proteome.
 SQ SEQUENCE 218 AA; 23476 MW; ECCCEA924F62D06A CRC64;

Query Match 22.2%; Score 201.5; DB 16; Length 218;
 Best Local Similarity 34.8%; Pred. No. 2.2e-08;
 Matches 47; Conservative 27; Mismatches 52; Indels 9; Gaps 3;

QY 53 PKQBPAPVVEKPAVESGTIIASIVDFDKYIEKESQDETLEIVQKAKENHMQVLE-G 111
 Db 78 PQAQTPAAKPVVSGSEKVTFAADALFDKAVLKPEGSKLDLVSKLGLTLEVVVIG 137

QY 112 NTDFGSGSEYNQALGVRTLSVKNALVIKGVKDKMKTISFGTK----PKAQKTR-- 165
 Db 138 HTDSFGSKYNDRLSVRAEAVRGYLVSKGIEANRVYTEGKRQLKVDPKCKGARKAQ 197

QY 166 --CYKENRRVDVKLM 178
 Db 198 IACQPNRRVEVV 212

RESULT 15
 Q8D2E3 PRELIMINARY; PRT; 170 AA.

AC Q8D2E3;
 DT 01-MAR-2003 (TremBLrel. 23, Created)
 DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
 DE Pal protein.
 GN PAL.
 OS Wigglesworthia brevipalpis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Wigglesworthia.
 OX NCBI_TaxID=164609;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22297718; PubMed=12219091;
 RA Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,
 RA Aksoy S.;
 RT "Genome sequence of the endocellular obligate symbiont of tsetse
 RT flies, Wigglesworthia glossinidia.";
 RL Nat. Genet. 32:402-407 (2002).
 DR EMBL; AB063522; BAC24557.1; -
 KW Complete proteome.
 SQ SEQUENCE 170 AA; 19249 MW; EBABE6068BF0F317 CRC64;

Query Match 22.1%; Score 200; DB 16; Length 170;
 Best Local Similarity 32.0%; Pred. No. 2.2e-08;
 Matches 55; Conservative 26; Mismatches 77; Indels 14; Gaps 4;

QY 3 RSSVFSFLVAFLLVAGCGSHMDNKTVAGDVSAKTQVATFVTEPAPEKEPKQAPVVE 62
 Db 8 KSVLS--IPLISMISCTSKQVKNIDEPVSSSENDSNVNIMPEGE-----LS 57



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:28:11 ; Search time 34.6806 Seconds
(without alignments)
878.747 Million cell updates/sec

Title: US-10-080-113-3

Perfect score: 974

Sequence: 1 MLEKFLKSLKQLFLGLGLV.....VEBIFKDSVNYGVKVPFIAM 192

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 159726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 19Jun03.*

```

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	970	92.5	192	AAW98389	H. pylori GHPO 38
2	932	95.7	192	AAW88597	H. pylori membrane
3	925	95.0	192	AAW55579	H. pylori ORF 06ep
4	800	82.1	160	AAW99884	Antigen 1 from clu
5	795	81.6	159	AAW98982	Protein encoded by
6	102	10.5	80	AAW55319	H. pylori ORF h2e
7	99	10.2	1279	AAW83047	S. epidermidis ope
8	99	10.2	3696	ABP40235	Staphylococcus epi
9	92.5	9.5	2349	AAO16359	Human translocated

10	92.5	9.5	2400	22	ABG20278	Novel human diago
11	92.5	9.5	2415	22	ABG20279	Novel human diago
12	91.5	9.4	450	22	ABG92607	Human protein sequ
13	91.5	9.4	450	23	ABF43910	FLJ10210 fis clone
14	87.5	9.0	394	23	ABG53347	Lactococcus lactis
15	84.5	8.7	527	12	AAAR14183	Product of clone I
16	84.5	8.7	527	14	AAAR38099	Schistosomula mans
17	84	8.6	1250	24	AAO26231	MDPT related human
18	84	8.6	6815	22	ABG66811	Drosophila melanog
19	83	8.5	472	22	ABG82129	S. epidermidis ope
20	83	8.5	491	23	ABF39882	Staphylococcus epi
21	83	8.5	835	19	AAW37882	BRCA1 modulator pr
22	83	8.5	835	20	AAW30150	Amino acid sequenc
23	83	8.5	1479	23	ABG09519	Human pregnancy zo
24	82	8.4	230	21	AAW18323	plasmodium falcipa
25	82	8.4	652	18	AAW18010	plasmodium falcipa
26	81.5	8.4	243	23	ABG49198	Listeria monocytog
27	81.5	8.4	1552	22	ABG71164	Drosophila melanog
28	81	8.3	424	22	ABG11569	Novel human diago
29	81	8.3	436	23	ABF30085	Streptococcus poly
30	81	8.3	1052	22	ABG15427	Novel human diago
31	81	8.3	1543	22	ABG69213	Drosophila melanog
32	81	8.3	1942	22	ABG15773	Novel human diago
33	81	8.3	2424	22	ABG13503	Novel human diago
34	80.5	8.3	2440	18	AAW20828	H. pylori cytoplas
35	80	8.2	735	20	AAW37551	Amino acid sequenc
36	80	8.2	1415	24	ABF81268	Arabidopsis thalia
37	80	8.2	1493	22	ABF72444	UGGT. Caenorhabdi
38	80	8.2	1881	23	ABF73809	Candida albicans e
39	80	8.2	1948	22	ABG21233	Novel human diago
40	79	8.1	356	23	ABG26464	Streptococcus poly
41	79	8.1	369	21	AAW37344	Arabidopsis thalia
42	79	8.1	391	21	AAW37343	Arabidopsis thalia
43	79	8.1	408	21	ABG37342	Arabidopsis thalia
44	78.5	8.1	249	23	ABU51175	Helicobacter pylor
45	78.5	8.1	274	23	ABG92690	Herbicidially activ

ALIGNMENTS

RESULT 1

AAW98389
ID AAW98389 standard; Protein; 192 AA.

XX AC AAW98389;

DT 31-MAR-1999 (first entry)

XX DE H. pylori GHPO 38 protein.

XX GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;

XX KW peptic ulcer disease.

XX OS Helicobacter pylori.

XX FN WO9843478-A1.

XX PD 08-OCT-1998.

XX PF 01-APR-1998; 98WO-US06371.

XX PR 29-JUL-1997; 97US-0902615.

XX PR 01-APR-1997; 97US-0833457.

XX PR 24-JUN-1997; 97US-0881227.

XX (HUMA-) HUMAN GENOME SCI INC.

XX PA (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

XX PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;

XX WPI; 1998-542293/46.

XX DR N-FSDE; AAX14108.

XX New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases
 XX
 PS Claim 8; Page 616-617; 2054pp; English.
 XX
 CC This sequence represents a Helicobacter pylori GPHO protein of the
 CC invention. The polypeptides can be used for preventing or treating
 CC Helicobacter infections, and gastroduodenal diseases associated with
 CC these infections, including acute, chronic, and atrophic gastritis, and
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 CC used for the production of antibodies. The products can also be used for
 CC detection and diagnosis.
 XX
 SQ Sequence 192 AA;
 Query Match 99.6%; Score 970; DB 19; Length 192;
 Best Local Similarity 99.5%; Pred. No. 3.2e-94;
 Matches 191; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLEKSFLLKSKQLFLCGLGVLMLQACTCPNTSQRNSFLQDVPYVWMLQNRSEYITQGVDSH 60
 DB 1 VLEKSFLLKSKQLFLCGLGVLMLQACTCPNTSQRNSFLQDVPYVWMLQNRSEYITQGVDSH 60
 QY 61 IVDGKTEIEIKIATKRAIRVAQNIYVHKLKAYLSKTNRIKOKITNEMFIQMTQPIYDS 120
 DB 61 IVDGKTEIEIKIATKRAIRVAQNIYVHKLKAYLSKTNRIKOKITNEMFIQMTQPIYDS 120
 QY 121 LMNVDRGLGIYINPNNEVFALVRARGFDKDALSEGHLKMSLDNQAVSILVAKVEEIFKDS 180
 DB 121 LMNVDRGLGIYINPNNEVFALVRARGFDKDALSEGHLKMSLDNQAVSILVAKVEEIFKDS 180
 QY 181 VNYGDVKVPIAM 192
 DB 181 VNYGDVKVPIAM 192
 RESULT 2
 AAY88597
 ID AAY88597 standard; Protein; 192 AA.
 AC AAY88597;
 XX
 DT 17-AUG-2000 (first entry)
 XX
 DE H. pylori membrane-related antigen protein sequence.
 XX
 KW Membrane-related antigen; Helicobacter pylori; vaccine; duodenal ulcer;
 KW gastritis; gastric ulcer; gastric cancer; gastric MALT lymphoma.
 XX
 OS Helicobacter pylori.
 XX
 PN JP2000083671-A.
 XX
 PD 28-MAR-2000.
 XX
 PF 11-SEP-1998; 98JP-0257343.
 XX
 PR 11-SEP-1998; 98JP-0257343.
 XX
 PA (SHIO) SHIONOGI & CO LTD.
 XX
 XX WPI; 2000-353335/31.
 DR N-PSDB; AAA30192.
 XX
 XX A vaccine composition for Helicobacter pylori infection -
 PT
 PS Claim 5; Page 13-14; 18pp; Japanese.
 XX
 CC This sequence represents a Helicobacter pylori membrane-related antigen
 CC amino acid sequence. The invention relates to a recombinant polypeptide
 CC having an amino acid sequence the same as that of H. pylori

CC membrane-related antigen of a molecular weight of approximately 20kDa
 CC (such as the protein represented by the present sequence). The invention
 CC also includes mutant H. pylori membrane-related antigen proteins. Also
 CC included in the invention is a peptide containing an immunogenic epitope
 CC of the protein, and a vaccine composition for inducing active immune
 CC response to H. pylori infection including the polypeptide and a
 CC pharmaceutically acceptable carrier. The polypeptide and the peptide are
 CC used for the treatment of duodenal ulcer, gastritis, gastric ulcer,
 CC gastric cancer or gastric MALT lymphoma.
 XX
 SQ Sequence 192 AA;
 Query Match 95.7%; Score 932; DB 21; Length 192;
 Best Local Similarity 94.8%; Pred. No. 3.4e-90;
 Matches 182; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MLEKSFLLKSKQLFLCGLGVLMLQACTCPNTSQRNSFLQDVPYVWMLQNRSEYITQGVDSH 60
 DB 1 MLEKSFLLKSKQLFLCGLGVLMLQACTCPNTSQRNSFLQDVPYVWMLQNRSEYITQGVDSH 60
 QY 61 IVDGKTEIEIKIATKRAIRVAQNIYVHKLKAYLSKTNRIKOKITNEMFIQMTQPIYDS 120
 DB 61 IVDGKTEIEIKIATKRAIRVAQNIYVHKLKAYLSKTNRIKOKITNEMFIQMTQPIYDS 120
 QY 121 LMNVDRGLGIYINPNNEVFALVRARGFDKDALSEGHLKMSLDNQAVSILVAKVEEIFKDS 180
 DB 121 LMNVDRGLGIYINPNNEVFALVRARGFDKDALSEGHLKMSLDNQAVSILVAKVEEIFKDS 180
 QY 181 VNYGDVKVPIAM 192
 DB 181 VNYGDVKVPIAM 192
 RESULT 3
 AAW55579
 ID AAW55579 standard; Protein; 192 AA.
 XX
 AC AAW55579;
 XX
 DT 07-JUL-1998 (first entry)
 XX
 DE H. pylori ORF 06ep30223_5271902_c1_106 secreted protein.
 XX
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 XX
 OS Helicobacter pylori.
 XX
 PN WO9737044-A1.
 XX
 PD 09-OCT-1997.
 XX
 PF 27-MAR-1997; 97WO-US052223.
 XX
 PR 06-DEC-1996; 96US-0761318.
 PR 29-MAR-1996; 96US-0625811.
 PR 02-APR-1996; 96US-0758731.
 PR 25-OCT-1996; 96US-0736905.
 PR 28-OCT-1996; 96US-0738859.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Alm RA, Smith D;
 XX
 DR WPI; 1997-503122/46.
 DR N-PSDB; AAV24988.
 XX
 XX Helicobacter pylori nucleic acid sequences and encoded
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
 PT infection and for diagnosis of H. pylori infection
 PS Claims 14, 94; Page 784; 1145pp; English.

XX This sequence is a H. pylori secreted protein. The protein may be used in
 CC a vaccine to prevent or treat H. pylori infection or to identify
 CC H. pylori polypeptide binding compounds, useful as potential H. pylori
 CC life cycle activators or inhibitors. The DNA and probes derived from it
 CC may be used for the identification of H. pylori in a sample and the
 CC diagnosis of H. pylori infection. Nucleic acid sequences complementary
 CC to the DNA act as antisense sequences and can be used to prevent the
 CC translation of H. pylori mRNA. Antibodies against the protein can be
 CC used in immunoassays to evaluate the abundance and distribution of
 CC H. pylori-specific antigens. The genomic sequence of H. pylori
 CC (ATCC 55679) was determined from overlapping contigs generated by
 CC mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions
 CC defined by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from H. pylori by PCR
 CC amplification for recombinant polypeptide production, e.g. in E. coli
 CC hosts.

XX Seq Sequence 192 AA;

Query Match 95.0%; Score 925; DB 18; Length 192;
 Best Local Similarity 93.8%; Pred. No. 1.9e-89;
 Matches 180; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

QY 1 MLEKSLKSKQLFLGGLVLMLOACTCPNTSORSNPFLOVPVWMLQNRSEYITQGVDSH 60
 DB 1 VLEKSLKSKQLVLCGLVLMLOACTCPNTSORSNPFLOVPVWMLQNRSEYITQGVDSH 60
 QY 61 IVDGKTEIEIKATKRATIRVAQNIVHKLKAYLSKTNRIKQKTNEMFIQMTPIYDS 120
 DB 61 IVDGKTEIEIKATKRATIRVAQNIVHKLKAYLSKTNRIKQKTNEMFIQMTPIYDS 120
 QY 121 LNMVRLGLIYNPNNEEVPALVRARGFQDKALSEGHLKMSLDNQAVSILVAKVEIFKDS 180
 DB 121 LNMVRLGLIYNPNNEEVPALVRARGFQDKALSEGHLKMSLDNQAVSILVAKVEIFKDS 180
 QY 181 VNYGDKVKPIAM 192
 DB 181 INYGVKVPPIAM 192

RESULT 4
 AA89884
 ID AA89884 standard; Protein; 160 AA.
 AC AA89884;
 XX

DT 20-MAR-2003 (updated)
 DT 18-FEB-1999 (first entry)

DE Antigen 1 from cluster 21a.

KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.

XX Helicobacter pylori.

XX WO9849314-A2.

XX 05-NOV-1998.

XX 25-APR-1998; 98WO-US08487.

XX 14-OCT-1997; 97US-0061958.

XX 25-APR-1997; 97US-0045107.

XX (GENE-) GENELABS TECHNOLOGIES INC.

XX Chow TP, Fry KE, Lim MY, McAttee CP;

XX WPI; 1999-009433/01.

XX - New Helicobacter pylori antigens and related nucleic acid sequences
 PT useful in serological diagnosis and protective vaccines, providing
 PT long-lasting immune response

PS Claim 1; Page 228-229; 402pp; English.

XX The present sequence represents a Helicobacter pylori antigenic protein
 CC that is characterised by immunoreactivity with H. pylori-positive
 CC antisera. The proteins are highly immunogenic and induce a long-lasting
 CC immune response that persists even after antimicrobial treatment. In
 CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
 CC highly sensitive and specific. The specification also describes 69
 CC previously unrecognised immunogenic cluster families. H. pylori antigens
 CC are used to detect H. pylori-specific antibodies, for diagnosing
 CC infection or to confirm eradication of infection, and in vaccines to
 CC protect against H. pylori infection and related diseases (gastritis,
 CC peptic ulcer, gastric adenocarcinoma/lymphoma).
 CC (Updated on 20-MAR-2003 to correct PF field.)

XX Seq Sequence 160 AA;

Query Match 82.1%; Score 800; DB 20; Length 160;
 Best Local Similarity 98.8%; Pred. No. 2.3e-76;
 Matches 158; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 33 RNSFLQDPVPMQLQNRSEYITQGVDSHIVDGKTEIEIKATKRATIRVAQNIVHKLKE 92
 DB 1 RNSFLQDPVPMQLQNRSEYITQGVDSHIVDGKTEIEIKATKRATIRVAQNIVHKLKE 60
 QY 93 AYLKSNRIKQKTNEMFIQMTPIYDSLNMVRLGLIYNPNNEEVPALVRARGFQDKAL 152
 DB 61 AYLKSNRIKQKTNEMFIQMTPIYDSLNMVRLGLIYNPNNEEVPALVRARGFQDKAL 120
 QY 153 SEGHLKMSLDNQAVSILVAKVEIFKDSVNYGDKVKPIAM 192
 DB 121 SEGHLKMSLDNQAVSILVAKVEIFKDSVNYGDKVKPIAM 160

RESULT 5
 AA89825
 ID AA89825 standard; Protein; 159 AA.
 AC AA89825;
 XX

DT 20-MAR-2003 (updated)
 DT 18-FEB-1999 (first entry)

DE Protein encoded by clone A22.

KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.

XX Helicobacter pylori.

XX WO9849314-A2.

XX 05-NOV-1998.

XX 25-APR-1998; 98WO-US08487.

XX 14-OCT-1997; 97US-0061958.

XX 25-APR-1997; 97US-0045107.

XX (GENE-) GENELABS TECHNOLOGIES INC.

XX Chow TP, Fry KE, Lim MY, McAttee CP;

XX WPI; 1999-009433/01.

XX N-PSDB; AAV90567.

PT New Helicobacter pylori antigens and related nucleic acid sequences
PT - useful in serological diagnosis and protective vaccines, providing
PT long-lasting immune response
XX
XX
XX Claim 15; Page 112; 402pp; English.
XX
XX The present sequence represents a Helicobacter pylori antigenic protein
CC that is characterised by immunoreactivity with H. pylori-positive
CC antisera. The proteins are highly immunogenic and induce a long-lasting
CC immune response that persists even after antimicrobial treatment. In
CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC highly sensitive and specific. The specification also describes 69
CC previously unrecognised immunogenic cluster families. H. pylori antigens
CC are used to detect H. pylori-specific antibodies, for diagnosing
CC infection or to confirm eradication of infection, and in vaccines to
CC protect against H. pylori infection and related diseases (gastritis,
CC peptic ulcer, gastric adenocarcinoma/lymphoma).
CC (Updated on 20-MAR-2003 to correct PF field.)
XX
XX Sequence 159 AA;
XX
XX Query Match 81.6%; Score 795; DB 20; Length 159;
XX Best Local Similarity 98.7%; Pred. No. 7.8e-76;
XX Matches 157; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 34 NSFLQDPVYMWLQNRSEYITQGVDSHIVDGGKTEIEKIATKRAIRVAQNIVHKLKEA 93
XX DB 1 NSFLQDPVYMWLQNRSEYITQGVDSHIVDGGKTEIEKIATKRAIRVAQNIVHKLKEA 60
XX
XX QY 94 YLSEKSNRIKQITNEMFIQMTPIYDSLMMVDRGLGIYINPNNEVPALVRARGFDKDAL 153
XX DB 61 YLSEKSNRIKQITNEMFIQMTPIYDSLMMVDRGLGIYINPNNEVPALVRARGFDKDAL 120
XX
XX QY 154 EGLHKNLDAQSVILVAKVEEIKFQSVNGYGVKVPIM 192
XX DB 121 EGLHKNLDAQSVILVAKVEEIKFQSVNGYGVKVPIM 159
XX
XX RESULT 6
XX AAW55319 standard; Protein; 80 AA.
XX AC AAW55319;
XX
XX DT 15-JUN-1998 (first entry)
XX DE H. pylori ORF hp2e10229orf4 protein.
XX DE Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
XX KW identification; binding compound; bacteria; life cycle; activator;
XX KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
XX
XX OS Helicobacter pylori.
XX
XX PN WO9737044-A1.
XX
XX PD 09-OCT-1997.
XX
XX PF 27-MAR-1997; 97WO-US05223.
XX
XX PR 06-DEC-1996; 96US-0761318.
XX PR 29-MAR-1996; 96US-0625811.
XX PR 02-APR-1996; 96US-0758731.
XX PR 25-OCT-1996; 96US-0736905.
XX PR 28-OCT-1996; 96US-0738853.
XX
XX PA (ASTR) ASTRA AB.
XX
XX PI Alm RA, Smith D;
XX
XX DR WPI; 1997-503122/46.
XX DR N-PSDB; AAV24728.
XX

PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX
XX PS Claim 14; Pages 546-547; 1145pp; English.
XX
XX This sequence is a H. pylori protein of unspecified function.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
XX
XX Sequence 80 AA;
XX
XX Query Match 10.5%; Score 102; DB 18; Length 80;
XX Best Local Similarity 87.5%; Pred. No. 0.0054;
XX Matches 21; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 MLEKSFLLSKQLFLCGLGLVLMQA 24
XX DB 19 VLEKSFLLSKQLFLCGLGLVLMQA 42
XX
XX RESULT 7
XX AAG83047
XX ID AAG83047 standard; Protein; 1279 AA.
XX AC AAG83047;
XX
XX DT 03-SEP-2001 (first entry)
XX DE S. epidermidis open reading frame protein sequence SEQ ID NO:3188.
XX KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX KW vaccination; endocarditis.
XX OS Staphylococcus epidermidis.
XX
XX PN WO200134809-A2.
XX
XX PD 17-MAY-2001.
XX
XX PF 09-NOV-2000; 2000WO-US30782.
XX
XX PR 09-NOV-1999; 99US-0164258.
XX
XX PA (GLAX) GLAXO GROUP LTD.
XX
XX PI Kimmerly WJ;
XX
XX DR WPI; 2001-316495/33.
XX DR N-PSDB; AAH53897.
XX
XX PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX
XX PS Claim 18; Page 842-843; 2188pp; English.
XX

CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AG81454 to AG83120, from *Staphylococcus epidermidis*.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC *S. epidermidis* polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects, and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of the
CC activity and therefore identify compounds that may be used for the
CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed *S. epidermidis* genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 1279 AA;
Query Match 10.2%; Score 99; DB 22; Length 1279;
Best Local Similarity 19.2%; Pred. No. 0.62;
Matches 37; Conservative 44; Mismatches 72; Indels 40; Gaps 7;
QY 30 TSQRNPLQDVPYWMQLQRSEY---ITQGVDSHIVDGK-----KTEEI 70
DB 721 TEKNTAQISDITLAQARNNINGANTVALVDENLEDGKQKLRIVLSTQTKQAKADIA 780
QY 71 EKATKATIRVAQNIHVHKLKEAVLSKTN-----RIKQKITNEMFIQMTQPIYDSLM 122
DB 781 QAIGQORSTIDONQATTEKQKALERLNOETNGVNDRIQALANQVTDKKNILETIR 940
QY 123 NVDRGLGIYINPNEEVFALVRARGFDKDALSEGLHKMSLDNQAVSILVAKVERIFKDSVN 182
DB 841 NVEPI-VIVPKANEI---IRKKAQETTLNQNDATLEEKQIAL--GKLEEVKNEALN 894
QY 183 Y-----GDVKV 188
DB 895 QVSOAHNSNNDVKI 907
RESULT 8
ABP40235
ID ABP40235 standard; Protein; 3696 AA.
XX
AC ABP40235;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5080.
XX
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.
XX
OS Staphylococcus epidermidis.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-0134001.
XX
PR 14-AUG-1997; 97US-055779P.
PR 08-NOV-1997; 97US-064964P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
PI
XX WPI; 2002-381255/41.

DR N-PSDB; ABN92780.
XX
PT Novel isolated nucleic acid encoding a *Staphylococcus epidermidis*
PT polypeptide, useful for diagnosing and treating bacterial infections -
PS Disclosure; SEQ ID 5080; 267pp; English.
XX
CC ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The *S. epidermidis* sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly *S. epidermidis* infections. The sequences can be used to
CC screen for compounds able to interfere with the *S. epidermidis* life
CC cycle or inhibit *S. epidermidis* infection.
CC N.B. the sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 3696 AA;
Query Match 10.2%; Score 99; DB 23; Length 3696;
Best Local Similarity 19.2%; Pred. No. 2.9; 72; Indels 40; Gaps 7;
Matches 37; Conservative 44; Mismatches 72; Indels 40; Gaps 7;
QY 30 TSQRNPLQDVPYWMQLQRSEY---ITQGVDSHIVDGK-----KTEEI 70
DB 2119 TEKNTAQISDITLAQARNNINGANTVALVDENLEDGKQKLRIVLSTQTKQAKADIA 2178
QY 71 EKATKATIRVAQNIHVHKLKEAVLSKTN-----RIKQKITNEMFIQMTQPIYDSLM 122
DB 2179 QAIGQORSTIDONQATTEKQKALERLNOETNGVNDRIQALANQVTDKKNILETIR 2238
QY 123 NVDRGLGIYINPNEEVFALVRARGFDKDALSEGLHKMSLDNQAVSILVAKVERIFKDSVN 182
DB 2239 NVEPI-VIVPKANEI---IRKKAQETTLNQNDATLEEKQIAL--GKLEEVKNEALN 2292
QY 183 Y-----GDVKV 188
DB 2293 QVSOAHNSNNDVKI 2305
RESULT 9
AAO16359
ID AAO16359 standard; Protein; 2349 AA.
XX
AC AAO16359;
XX
DT 03-APR-2003 (first entry)
XX
DE Human translocated promoter region (TPR) protein, SEQ ID No 7.
XX
KW Human; p53 pathway modulating agent; p53-associated disorder;
KW translocated promoter region; TPR; cancer.
XX
OS Homo sapiens.
XX
PN WO2002099050-A2.
XX
PD 12-DEC-2002.
XX
PF 03-JUN-2002; 2002WO-US17425.
XX
PR 05-JUN-2001; 2001US-296076P.
PR 10-OCT-2001; 2001US-328605P.
PR 15-FEB-2002; 2002US-357253P.
XX
PA (EXEL-) EXELIXIS INC.
XX
XX Friedman L, Plowman GD, Belvin M, Francis-lang H;
XX
XX WPI; 2003-167335/16.

PT Identifying candidate p53 pathway-modulating agents, useful as
 PT therapeutic targets for disorders associated with defective p53
 PT function, comprises screening for agents that modulate the activity of
 PT translocated promoter region polypeptides -
 XX
 XX Claim 13; Page 84-94; 94pp; English.
 XX
 XX The invention comprises a method for identifying a candidate p53 pathway
 CC modulating agent. The method involves providing an assay system
 CC comprising a purified translocated promoter region (TPR) protein or
 CC nucleic acid. The method of the invention is useful for identifying
 CC candidate p53 pathway modulating agents for use as therapeutic targets
 CC for disorders associated with defective p53 function (e.g. cancer). The
 CC TPR proteins and nucleic acids are useful for identifying agents that
 CC modulate TPR function. The present amino acid sequence represents a human
 CC TPR protein.
 XX
 SQ Sequence 2349 AA;
 Query Match 9.5%; Score 92.5; DB 24; Length 2349;
 Best Local Similarity 24.1%; Pred. No. 7.3;
 Matches 48; Conservative 38; Mismatches 74; Indels 39; Gaps 10;
 QY 17 LGVLMQACTCPTNSQNSFLQ-----DVPYMLQNRSEYITQGV-----S 58
 Db 1302 LDILPQEAANA-ELSEKSGMLQAEKLLLEEDVVRKARNO-HLVSQKQDPDTEYRKLLS 1359
 QY 59 SHIVDGKK-----TEEI-----EKIATKATIRVAQNIHVHKLKAEVLSKTNRIKQITNEMF 110
 Db 1360 EKEVHTKRIQQLTHEIGELKAEIARSNASLTNNQNLQSLKED-LNKVTEKETIQKDL 1418
 QY 111 IQM--TQPIYDSLMNVDRLGIYINPNNEEVFALVRARGFDKDALSGHLKMSLDNQAVSI 168
 Db 1419 AKIIDIQEKVXTITQVKIGRRYKTYQYELKA-----QDDKVMETSAQSSGDHQEHV 1471
 QY 169 LVAKVEEIFKDSVNYGDVK 187
 Db 1472 SVQEMQEL-KETLNQAEATK 1489
 RESULT 10
 ABG20278
 ID ABG20278 standard; Protein; 2400 AA.
 AC
 XX ABG20278;
 XX 18-FEB-2002 (first entry)
 XX
 XX Novel human diagnostic protein #20269.
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 XX WO200175067-A2.
 XX
 XX 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US08631.
 XX
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Drmanac RT, Liu C, Tang YT;
 XX
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS84465.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 50637; 103pp; English.
 XX
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG03077 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2400 AA;
 Query Match 9.5%; Score 92.5; DB 22; Length 2400;
 Best Local Similarity 24.1%; Pred. No. 7.6;
 Matches 48; Conservative 38; Mismatches 74; Indels 39; Gaps 10;
 QY 17 LGVLMQACTCPTNSQNSFLQ-----DVPYMLQNRSEYITQGV-----S 58
 Db 1331 LDILPQEAANA-ELSEKSGMLQAEKLLLEEDVVRKARNO-HLVSQKQDPDTEYRKLLS 1388
 QY 59 SHIVDGKK-----TEEI-----EKIATKATIRVAQNIHVHKLKAEVLSKTNRIKQITNEMF 110
 Db 1389 EKEVHTKRIQQLTHEIGELKAEIARSNASLTNNQNLQSLKED-LNKVTEKETIQKDL 1447
 QY 111 IQM--TQPIYDSLMNVDRLGIYINPNNEEVFALVRARGFDKDALSGHLKMSLDNQAVSI 168
 Db 1448 AKIIDIQEKVXTITQVKIGRRYKTYQYELKA-----QDDKVMETSAQSSGDHQEHV 1500
 QY 169 LVAKVEEIFKDSVNYGDVK 187
 Db 1501 SVQEMQEL-KETLNQAEATK 1518
 RESULT 11
 ABG20279
 ID ABG20279 standard; Protein; 2415 AA.
 AC
 XX ABG20279;
 XX 18-FEB-2002 (first entry)
 XX
 XX Novel human diagnostic protein #20270.
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 XX WO200175067-A2.
 XX
 XX 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US08631.
 XX
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.

XX (HYSZ-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 XX N-PSDB; AAS84466.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 20; SEQ ID No 50638; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAG0010-AAG0377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 2415 AA;

Query Match 9.5%; Score 92.5; DB 22; Length 2415;
 Best Local Similarity 24.1%; Pred. No. 7.6;
 Matches 48; Conservative 38; Mismatches 74; Indels 39; Gaps 10;
 QY 17 LGVLMQACTCPNTSQNSFLQ-----DVPYMLQNRSEYITQGV-----S 58
 Db 1364 LDILPQEAANA-ELSEKSGMLQAEKULEEDVKEWKARNQ-HLVSQCKDPDTEYRKLLS 1421
 QY 59 SHIVDCK-----TDEI-----EKIATKATIRVAQNIHVHKLKEAVLSKTNRIKQINEMF 110
 Db 1422 EKEVHKRIQQLTEIGSLKAEIARSNASLTNNQNLQSLKED-LNKVTRKETIQKLD 1480
 QY 111 IQM--TQPIYDLSMNVDRIGTYIINPNNEEVPALVRARGFDKDALSEGLHKMSLDNQAVSI 168
 Db 1481 AKIIDIQEKVKTITQVKKIGRRYKTOYEELKA-----QQDKVMTSAGSSGDHQEQHV 1533
 QY 169 LVAKVEEIPKSVNYGVVK 187
 Db 1534 SVQEMQEL-KETLQAETK 1551

RESULT 12
 AAB92607
 ID AAB92607 standard; Protein; 450. AA.
 AC AAB92607;
 XX 26-JUN-2001 (first entry)
 DT Human protein sequence SEQ ID NO:10869.
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 KW Homo sapiens.
 XX OS

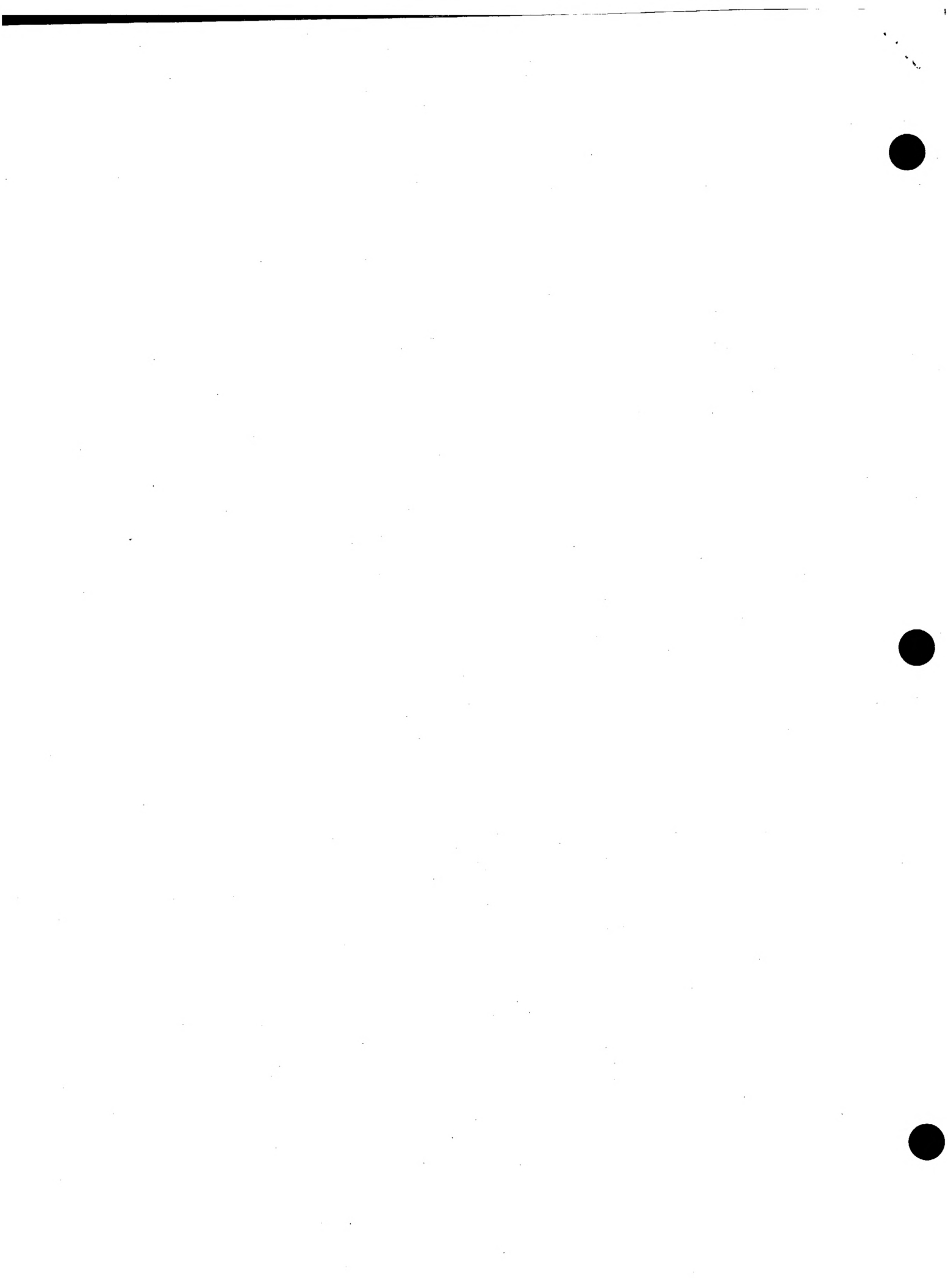
XX EP1074617-A2.
 XX PD 07-FEB-2001.
 XX 28-JUL-2000; 2000EP-0116126.
 XX 29-JUL-1999; 99JP-0248036.
 XX 27-AUG-1999; 99JP-0300253.
 XX 11-JAN-2000; 2000JP-0118776.
 XX 02-MAY-2000; 2000JP-0183767.
 XX 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI; 2001-318749/34.
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX Claim 8; SEQ ID 10869; 2537pp + CD ROM; English.
 XX The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX Sequence 450 AA;

Query Match 9.4%; Score 91.5; DB 22; Length 450;
 Best Local Similarity 23.9%; Pred. No. 0.85;
 Matches 48; Conservative 27; Mismatches 61; Indels 65; Gaps 10;
 QY 20 LMLQACTCPNTSQNSFLQDVPYMLQNRSEYITQGVDS--HIVDGKTEI--EKIAT 75
 Db 1 MVQAAVAPNRSQ--LLKIPYGLSLRRSVERMTEGRCQVHLLDDRKLLELVQPKLA 58
 QY 76 KBATIRVAQNIHVHKLKE-----AY 94
 Db 59 KELLDLVASHFNLKEKEYFGIATDTGHLNQLDRRVLEHDFPKSGFVVLFCVRFY 118
 QY 95 LSKTNRIKQKITNEM-FIQMTQPIYDLSMNVDRIGTYIINPNNEEVPALV-----RARG-- 146
 Db 119 IESISYLNKDNTATTELPFLNAKSCITYKELIDVD-----SEWVFELASYILQAKGDF 169
 QY 147 FDKDALSEGLHKM-SLDNQAV 166
 Db 170 SSNEVVRSDLKUPALPTQAL 190

Db	119	IESISYLNKDNTATIEFLFNNAKSCIYKELIDVD-----SEVVFELASYILOEAKGDF	166
Qy	147	FDKDALSEGLHKM-SLDNQAV	166
Db	170	SSNEVVRSDLKLKLPALPTQAL	190
RESULT 14			
ABBS3347			
ID	ABBS3347	standard; Protein; 394 AA.	
XX			
AC	ABBS3347		
DT	16-WAY-2002	(first entry)	
XX			
DE	Lactococcus lactis	protein ps123.	
XX			
KW	Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.		
XX			
OS	Lactococcus lactis	IL1403.	
XX			
PN	FR2807446-A1.		
XX			
PD	12-OCT-2001.		
XX			
PF	11-APR-2000; 2000FR-0004630.		
XX			
PR	11-APR-2000; 2000FR-0004630.		
XX			
PA	(INRG) INRA INST NAT RECH AGRONOMIQUE.		
XX			
PI	Bolotine A, Sorokine A, Renault P, Ehrlich SD;		
XX			
XX	WPI; 2002-043418/06.		
XX			
PT	New nucleotide sequence useful in the identification or Lactococcus		
PT	lactis and related species -		
XX			
PS	Claim 6; SEQ ID No 49; 2504pp; French.		
XX			
CC	The present invention is related to a Lactococcus lactis nucleotide		
CC	sequence (AB90521) and related proteins (AB953300-AB955621). The		
CC	nucleic acid sequence is useful in the detection and/or amplification of		
CC	nucleic acid sequence, particularly to identify Lactococcus lactis or		
CC	related species. The proteins of the invention are useful for the		
CC	biogenesis or biodegradation of a composition of interest. The		
CC	invention helps research in lactic bacteria, particularly useful in the		
CC	production of yogurt and cheese.		
CC	Note: The sequence data for this patent is based on equivalent patent		
CC	WO200177334 (published 18-OCT-2001) which is available in electronic		
CC	format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.		
XX			
SQ	Sequence 394 AA;		
Query Match			
Best Local Similarity	9.0%; Score 87.5; DB 23; Length 394;		
Matches	33; Conservative 19; Mismatches 37; Indels 23; Gaps 6;		
Qy	72	KIATKRTIRVAON-----IVHKLKZAYLSKTR-IKKKTNMEFIQTOIYDSLMNVD	125
Db	81	KLGVKPRVQIMQTRNLNNVILPALGETRIDKNSILQOIVNQWMINASQPLN-----	133
Qy	126	RLGIYINPNB-----EVFALVRARGDKDALSEGLHKMSLDNQAVSILVAKV	173
Db	134	--GAYHPRKGGKDFKIYFIVERIF-KHALSLGLVK---DNPCVNVIVPKV	179
RESULT 15			
AAR14183			
ID	AAR14183	standard; Protein; 527 AA.	
XX			
AC	AAR14183		
XX			

	DT	25-MAR-2003	(updated)	
	XX	03-JAN-1992	(first entry)	
	DE	Product of clone Irv-5.		
	XX			
	KW	Epitope; bilharziasis; antibodies; schistosomula; vaccine;		
	XX	immunogen; anti-Irv; fusion.		
	OS	Schistosoma mansoni.		
	XX			
	FH	Key	Location/Qualifiers	
	FT	Domain	121..572	
	FT	/label= alpha helix		
	FT	Modified-site	160	
	FT	/label= N-glycos_site		
	FT	Modified-site	218	
	FT	/label= N-glycos_site		
	XX	US0501254-A.		
	PN			
	XX	24-SEP-1991.		
	PD			
	XX	30-SEP-1988; 88US-0252075.		
	PFF			
	PR	30-SEP-1988; 88US-0252075.		
	XX	(UYJO) UNIV JOHNS HOPKINS.		
	PPA			
	XX	Strand M;		
	PI			
	XX	WPI; 1991-303C03/41.		
	DR	N-PADB; AAQ13982.		
	DR			
	XX			
	PPT	Polypeptide(s) specific for epitope(s) of Schistosoma mansoni -		
	PT	used as vaccines against bilharziasis.		
	XX			
	PS	Claim 5; Fig 1B; 1Opp; English.		
	CC	The clone was obt'd. by screening an adult worm cDNA library in		
	CCC	lambda-gt11 with anti-Irv sera raised against a subset of adult worm		
	CCC	antigens. The sequence is thought to at least partially encode a		
	CCC	schiostome myosin. It has 48% identity with the myosin heavy		
	CCC	chain of C. elegans. The DNA can be used to produce a fusion protein		
	CCC	with beta galactosidase sequences. The product can be used as an		
	CCC	immunogen to produce vaccines against bilharziasis.		
	CCC	(Updated on 25-MAR-2003 to correct PA field.)		
	XX			
	SS	Sequence	527 AA;	
	SQ			
		Query Match	8.7%; Score 84.5; DB 12; Length 527;	
		Best Local Similarity	22.5%; Pred. No. 5.9;	
		Matches	38; Conservative 35; Mismatches 45; Indels 51; Gaps 8;	
QY		32 QRN-----SFLQDVPVWLQNRSBYITGGVDGSHIVDGKKTEIEIKATKRATIRVAQNI V 87		
Dd		98 QRNIRKYVLNLNFWPRLYT-----XVKPLMLNTAQEEEMKADEL 139		
QY		88 HKLEAYLSKTNR IKOXI-----TNMEFIQTPIYDSLNVDLGLIYNPNNEE 137		
Dd		140 AKLNEY-EKEUKLEELFQNTVTVIQQKNDLFLQL-QTEQSLADAE-----EK 187		
QY		138 VFALVRAGFDKDALS EGLHKMS-LDNQA VSI-----LVAKVEIFDK 179		
Dd		188 ITKVLOGNDMEORIKLEERLAEDDOANLT EVKKKSABIELKKD 236		

Search completed: December 10, 2003, 18:33:45
Job time : 35.6806 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:31:22 ; Search time 12.0628 Seconds
(without alignments)
673.448 Million cell updates/sec

Title: US-10-080-113-3

Perfect score: 974

Sequence: 1 MLEKSFLLKSLFLCGLGLV.....VEEIPKDSVNYGDKVPIAM 192

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgm2_6/ptodata/1/iaa/5A COMB.pcp.*
- 2: /cgm2_6/ptodata/1/iaa/5B COMB.pcp.*
- 3: /cgm2_6/ptodata/1/iaa/6A COMB.pcp.*
- 4: /cgm2_6/ptodata/1/iaa/6B COMB.pcp.*
- 5: /cgm2_6/ptodata/1/iaa/PTCUS COMB.pcp.*
- 6: /cgm2_6/ptodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99	10.2	3696	4	US-09-134-001C-5080
2	83	8.5	491	4	US-09-134-001C-4727
3	83	8.5	835	2	US-08-968-751-4
4	82	8.4	652	4	US-08-559-8968-2
5	81	8.3	516	4	US-09-107-532A-7156
6	80	8.2	1493	4	US-09-376-330-16
7	78.5	8.1	385	4	US-09-134-001C-5166
8	78	8.0	1189	1	US-08-602-737-2
9	78	8.0	1189	3	US-08-001-982-2
10	77.5	8.0	131	4	US-08-328-352-6148
11	77.5	8.0	10182	4	US-09-134-001C-3159
12	76	7.8	222	2	US-08-933-750C-41
13	76	7.8	222	3	US-09-234-613-41
14	76	7.8	248	4	US-09-134-001C-3211
15	76	7.8	276	4	US-09-328-352-6056
16	76	7.8	325	4	US-09-451-325-141
17	76	7.8	336	4	US-09-252-991A-2121
18	75	7.7	394	4	US-08-195-705-2
19	75	7.7	2639	3	US-09-080-983-3
20	74.5	7.6	433	4	US-09-345-236B-43
21	74.5	7.6	843	4	US-09-328-352-5294
22	74	7.6	508	4	US-09-126-420A-25
23	73.5	7.5	380	4	US-09-134-001C-3830
24	73.5	7.5	443	4	US-09-252-991A-27545
25	73	7.5	208	4	US-09-328-352-5545
26	73	7.5	614	4	US-09-134-001C-3061
27	73	7.5	767	4	US-09-252-991A-31198

ALIGNMENTS

RESULT 1

US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GFC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,564
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match 10.2%; Score 99; DB 4; Length 3696;

Best Local Similarity 19.2%; Pred. No. 0.75; Mismatches 44; Indels 72; Gaps 7;

Matches 37; Conservative 44; Mismatches 72; Indels 40; Gaps 7;

QY 30 TSORNSPLQDPVPMQLQNRSEY---ITQGVDSHIVDGH-----KTEEI 70

Db 2119 TEENKTAIQSIDDTLAQRNNINGANTNALVDENLEDGKQLQRIVLSTQTKTOAKADIA 2176

QY 71 EKATKRAATIRVAQNIHVHKLKEAYLSKTN-----RIKQKITMEMFTQTPYDSLM 122

Db 2179 QATGQKSTIDQNCNATTEKQALERLNOETNGVNDRIQALANQNTVDEKKNILETR 2238

QY 123 NVDRGLGIYINPNEEYFVALVRGFDKDALSEGLHKNSLDNQAVSTILVAKVEEIFKDSVN 182

Db 2239 NVEPI-VIVKPKANEI--IRKKAASQTTLNQNDATLEEKQIAL--GKLESEKNEALN 2292

QY 183 Y-----GVKV 188

Db 2293 QVSOAHNSNDVKI 2305

RESULT 2

US-09-134-001C-4727
; Sequence 4727, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4727
LENGTH: 491
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4727

Query Match 8.5%; Score 83; DB 4; Length 491;
Best Local Similarity 22.2%; Pred. No. 2.1;
Matches 39; Conservative 37; Mismatches 66; Indels 34; Gaps 9;

QY 19 VLMQACTCPNTSQRNSFLQDVPYVWLNQNRSEYITQGVDS--HIV--DGKTEIEKI 73
DB 191 IIDMSCLIQD-RQHQEVNVRVWLNELNISYNEKTKGLIRHLVVRGTHDEWVI 249

QY 74 ATKRATIRVAQNIHVHLKEAYLSKTNRIKQKITNEMFIQMTQPIYDSLNVDRGLIYN 132
DB 250 FVTNGATFKQSELLVKNLKKEPPNITS-IKQINNSH-----SNVIMGROSMTLY-- 298

QY 133 PNNEEVFALVRGDFK--DALSEGLHKMSLDNQAVSILVAKVEIFKDSVNYGVDV 186
DB 299 -----GDKIEDQSEVTHYIS-DLSFYQINSSQTEKYQQALNVAQL 340

RESULT 3

US-08-968-751-4
Sequence 4, Application US/08968751
Patent No. 5948643
GENERAL INFORMATION:
APPLICANT: Rubinfeld, Bonnie
APPLICANT: Polakis, Paul G.
APPLICANT: Ligenfelter, Carol
APPLICANT: Vuong, Terilyn T.
TITLE OF INVENTION: MODULATORS OF BRCA1 ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: ONYX Pharmaceuticals, Inc.
STREET: 3031 Research Drive
CITY: Richmond
STATE: CA
COUNTRY: USA
ZIP: 94806
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,751
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Giotta, Gregory
REGISTRATION NUMBER: 32,028
REFERENCE/DOCKET NUMBER: ONYX1024 GG
TELEPHONE: (510) 262-8710
TELEFAX: (510) 222-9758
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 835 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-968-751-4

RESULT 4

US-08-559-896B-2
Sequence 2, Application US/08559896B
Patent No. 6310046
GENERAL INFORMATION:
APPLICANT: Patrick E. Duffy
APPLICANT: Christian F. Ockenhouse
TITLE OF INVENTION: SEQUESTRAIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: John Moran
STREET: USA MEMC - MCMR-JA
CITY: FORT DETRICK, FREDERICK
STATE: MARYLAND
COUNTRY: USA
ZIP: 21702-5012
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.5
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/559,896B
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Moran, John
REGISTRATION NUMBER: 26,313
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 619-2065
TELEFAX: (301) 619-7714
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 652 amino acids
TYPE: amino acid
TOPOLOGY: Linear
US-08-559-896B-2

Query Match 8.4%; Score 82; DB 4; Length 652;
Best Local Similarity 22.4%; Pred. No. 4;
Matches 43; Conservative 20; Mismatches 71; Indels 58; Gaps 7;

QY 47 NRSEYITQGVDSHIVDGKTE-----EIEKIATKRATIRVAQNI-----VHK----- 89
DB 29 NISANTNTSTKESHILDESKELEFYRDELDMGKBEISTYFKGNIDKXSLDEPHKILLE 88
QY 90 -----LKEAYLSKTNRIKQKITNEMFIQMTQPIYDSLNVDRGLIYNPNNEEVFAL 141


```

; SEQ ID NO 5166
; LENGTH: 385
; TYPE: PRI
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5166

Query Match      8.1%; Score 78.5; DB 4; Length 385;
Best Local Similarity 22.8%; Pred. No. 4.4;
Matches 49; Conservative 27; Mismatches 78; Indels 61; Gaps 9;

QY 11 QLFCLGGLVLMQACITPNTSQNSFLQDVPYMWLQNRSEYITQGVDSH-----60
DB 145 QLSVKVLEKIIAQNFVSTETRPVLTVGNWLIQD-NELICTATDGHRAVRKLQED 202
QY 61 -----IVDGKTEETEKI-----ATKRAITIRVAQ-NIVHKLKEAYLSKTRN 100
DB 203 ESENKWIIPGKALSELNKLMSDSDSDIDIFFASNOVLPRVGNINFIKRLLEGHYDPDTR 262
QY 101 -----IKQITNEFIQMTQPIYDSLNMVDPRLGIYINPNNEEVFALVRARGPKDALS 153
DB 263 LFPENYEIKLGNDFYHA-----IDRASL-----LAREGGNNVIKLS 301
QY 154 EGLHNSLDNQAVSILVAKVEEIFFKDSVNYGDKV 188
DB 302 TGNELVELSSTSPGIVTK-EEVNANDVEGNLKI 335

RESULT 8
US-08-602-737-2
; Sequence 2, Application US/08602737
; Patent No. 5736131
; GENERAL INFORMATION:
; APPLICANT: Bosch, Hendrick J.
; APPLICANT: Stiekema, Willem J.
; TITLE OF INVENTION: Hybrid Toxin
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SANDOZ AGRO. INC
; STREET: 975 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: CA 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,737
; FILING DATE: 21-FEB-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Marcus-Wyner, Lynn
; REGISTRATION NUMBER: 34,869
; REFERENCE/DOCKET NUMBER: 130-4080
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-354-3588
; TELEFAX: 415-857-1125
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1189 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-602-737-2

Query Match      8.0%; Score 78; DB 1; Length 1189;
Best Local Similarity 19.8%; Pred. No. 24;
Matches 41; Conservative 38; Mismatches 46; Indels 82; Gaps 9;

QY 32 QRNSFLQDVPYMWLQNRSEYITQGVDSHIVDGKTEETEKIATKRAITIRVAQNVHKLK 91
DB 2 EENNQCIPYNCLSNPEEVLQD-----ERISTGNSSIDISLSLVQFLV 46
QY 92 -----BAYLSKTRNRIKQITNEMFIQMTQPIYDSLMMVD 125
DB 47 SNFVPGGGLVGLIDFVWGIQVPSQWDAFLVQI-----EQLINERIAEFARNA--AIANLE 100
QY 126 RLQ-----IYI-----NPNNEEVFALVRARGPKDALSSEGLHK-----158
DB 101 GLGNFNLIYVAFKWEEDENNPE-----TETRVIDRFRILDGLLERDIPSFRISSGFEVPL 156
QY 159 MSLDNQAVSILVAKVEEIFFKDSVNYGD 185
DB 157 LSVYAQAANLELA-----ILRDSVIFGE 179

RESULT 9
US-09-001-982-2
; Sequence 2, Application US/09001982
; Patent No. 6204246
; GENERAL INFORMATION:
; APPLICANT: Bosch, Hendrick J.
; APPLICANT: Stiekema, Willem J.
; TITLE OF INVENTION: Hybrid Toxin
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 620424artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/001,982
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/602,737
; FILING DATE: 21-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: 130-4080/PCT/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1189 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-001-982-2

Query Match      8.0%; Score 78; DB 3; Length 1189;
Best Local Similarity 19.8%; Pred. No. 24;
Matches 41; Conservative 38; Mismatches 46; Indels 82; Gaps 9;

QY 32 QRNSFLQDVPYMWLQNRSEYITQGVDSHIVDGKTEETEKIATKRAITIRVAQNVHKLK 91
DB 2 EENNQCIPYNCLSNPEEVLQD-----ERISTGNSSIDISLSLVQFLV 46
QY 92 -----BAYLSKTRNRIKQITNEMFIQMTQPIYDSLMMVD 125
DB 47 SNFVPGGGLVGLIDFVWGIQVPSQWDAFLVQI-----EQLINERIAEFARNA--AIANLE 100
QY 126 RLQ-----IYI-----NPNNEEVFALVRARGPKDALSSEGLHK-----158
DB 101 GLGNFNLIYVAFKWEEDENNPE-----TETRVIDRFRILDGLLERDIPSFRISSGFEVPL 156

```

QY 159 MSLDNQAVSILWAKVEEIFKDSVNYGD 185
: : | : : | : : | : :
DB 157 LSVYAQAANHLA----LRDSVIEGE 179

RESULT 12
US-08-933-750C-41
; Sequence 41, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Cosley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA

ADDRESS: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750C
FILING DATE: September 23, 1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0356 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166
 TELEX:
 INFORMATION FOR SEQ ID NO: 41:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 222 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: SININOT01
 CLONE: 2184712
 US-08-933-750C-41

Matches	47	Conservative	23	Mismatches	68	Indels	64	Gaps	10
Qy	3	EXPSLKSQIUFGLGLVLMQLQACTCPTNSQRSNLSFLQDVPYMWLQNRSEYITQGVDSHIV	62						
Db	30	EKILIK-KQBFLEQKIQEQLQAKKYGTKNKAALQ-----ALRRKKRPEQQAQT----	79						
Qy	63	DKG-----KTEIEKIAIKRATIRVAQNIVHKUKEAY-----LQKTNRIKQKITNEMFIQ	112						
Db	80	DGLSTLFEQREAIENATNAEVLRTMELAAQSKMKAYQDMIDKVDLMTMDITEQQ--E	137						
Qy	113	MTQPIYDSLMMVDRUGIYINPNNEEVPALVMARGF-----DKDAASEGLHKWGLDQAVSI	168						

Db 138 VAQQLSD-----AISRPMGFGDDVDEDELEEE----- 168

Qy 169 LVAKVEEIFKDSVNYGD-----VKVP 189
 Db 169 ---EQBELAQELINVGDKKEEPSVKLP 192

RESULT 13
 US-09-234-613-41
 ; Sequence 41, Application US/09234613
 ; Patent No. 6132973
 ; GENERAL INFORMATION:
 ; APPLICANT: Lal, Preeti
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Bandman, Olga
 ; APPLICANT: Shah, Purvi
 ; APPLICANT: Au-Young, Janice
 ; APPLICANT: Yue, Henry
 ; APPLICANT: Guegler, Karl J.
 ; APPLICANT: Corley, Neil C.
 ; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
 ; NUMBER OF SEQUENCES: 98
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/234,613
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/933,750
 ; FILING DATE: September 23, 1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0356 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 41:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 222 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: SININOT01
 ; CLONE: 2184712
 ; US-09-234-613-41

Query Match 7.8%; Score 76; DB 3; Length 222;
 Best Local Similarity 22.7%; Pred. No. 3.7;
 Matches 47; Conservative 28; Mismatches 69; Indels 64; Gaps 10;
 Qy 3 EXSFLSKQFLICGLGLVLMQACTCPTNSQNSFLQDVPYVWMLQNRSEYITQGVDSHIV 62
 Db 30 EXLIK-KQEFLEQKIQQLQSLQAKKYGTNKPAAQ-----ALRRKKRFFQQLAQT---- 79
 Qy 63 DGK-----KTEIEKIATKTRATIRVAQNI VHKLKAY-----LSKTNRIKOKITNEMFIQ 112
 Db 80 DGLTSLTEFOREAIENATINAEVLRTMELAAQSMKAYQDMIDIKVDELMTDITEQQ--E 137

Qy 113 MTOPIYDSLMMVDRIGIYINPNNEEVALYRARGF---DKDALSGLHKMSLDNQAVSI 168
 Db 138 VAQQLSD-----AISRPMGFGDDVDEDELEEE----- 168
 Qy 169 LVAKVEEIFKDSVNYGD-----VKVP 189
 Db 169 ---EQBELAQELINVGDKKEEPSVKLP 192

RESULT 14
 US-09-134-001C-3211
 ; Sequence 3211, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 3211
 ; LENGTH: 248
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-3211

Query Match 7.8%; Score 76; DB 4; Length 248;
 Best Local Similarity 18.7%; Pred. No. 4.4;
 Matches 32; Conservative 27; Mismatches 60; Indels 52; Gaps 6;
 Qy 10 KOLFCLGLGLVLMQACTCPTNSQNSFLQDVPYVWMLQNRSEYITQGVDSHIVG---KK 66
 Db 5 KNIIVIKLGGIAIE-----NLNDAFIQQINAHMLENKIIIVHG--GGQVISNLLTKN 55
 Qy 67 TEEIEKIATKTRATIR---VAQNTVHKLKEA----- 93
 Db 56 NHTSIKIDGKRVTAKNLPIIYDALINIVGHQLLERKESNLEFFQFKKIKELVSAEFL 115
 Qy 94 -----YLSKTNRIKOKITNEMFIQMTQPIYDSLMVDRIGIYINPNNEEV 138
 Db 116 NKNIYGVYGVKYEINTMLEKMLSRDIPIITSL-GVNEQGEYLVNADHL 165

RESULT 15
 US-09-328-352-6056
 ; Sequence 6056, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 6056
 ; LENGTH: 276
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 ; US-09-328-352-6056

Query Match 7.8%; Score 76; DB 4; Length 276;
 Best Local Similarity 17.6%; Pred. No. 5.1;
 Matches 36; Conservative 37; Mismatches 67; Indels 64; Gaps 5;
 Qy 26 TCPNTSQNSFLQDVPYVWMLQNRSEYITQGVDSHIV-----GKKTTEEIEKIATKTRATI 80
 Db 56 TPESVDARAFLOGGYQPLQQAQVHLLKQKAKWLDICGCGGYTSAQQVVEQICIGV 115

81	QY	RVAQNIYVHLKEAVL	SKTNIETAIKQI	TINEMFIQNTORIYD	SLMNV-----	124
		:	:	:	:	
		:	:	:	:	
116	DB	DIANNV-----	QRAKNKDVHV	VVTGTALP	VIDQSDMVCISL	167
		:	:	:	:	
		:	:	:	:	
125	QY	---DLGIIYNNEE	VFALVR	AGPKD	KAL	157
		:	:	:	:	
		:	:	:	:	
168	DB	LKDGYLIVVT	PATDTHIYAM-----	REALFEQVNP	HTPQFVEQLDFELK	220
		:	:	:	:	
		:	:	:	:	
158	QY	---KMSLDNO	QVASILVAK	VEE	IFKOS	180
		:	:	:	:	
		:	:	:	:	
221	DB	APLVLDQ	QALKNLIAM	TFAYKAS	244	
		:	:	:	:	
		:	:	:	:	

Search completed: December 10, 2003, 18:38:38
Job time : 13.0628 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:33:52 / Search time 23.623 Seconds
(without alignments)
1511.612 Million cell updates/sec

Title: US-10-080-113-3

Perfect score: 974

Sequence: 1 MLEKSFUKSKOLFCLGVL.....VERIFKDSVNYGVKVIPIAM 192

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09D_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	974	100.0	192	12	US-10-080-113-3
2	970	99.6	192	12	US-09-895-913A-8
3	82	8.4	652	9	US-09-351-794A-2
4	80	8.2	1881	12	US-10-032-585-7646
5	78.5	8.1	337	12	US-10-097-111-295
6	78.5	8.1	470	10	US-09-881-752A-174
7	78	8.0	1189	12	US-10-102-469-20
8	77.5	8.0	427	15	US-10-156-761-12081
9	77.5	8.0	856	9	US-09-815-242-11310
10	77.5	8.0	863	9	US-09-796-100-2
11	77.5	8.0	863	15	US-10-163-547-7
12	76.5	7.9	779	11	US-09-823-187-94
13	76	7.8	222	9	US-09-840-787-41
14	76	7.8	313	9	US-09-815-242-12014
15	76	7.8	325	15	US-10-012-542-141

Sequence 440, Appli
Sequence 7, Appli
Sequence 7, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 2, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 11306, A
Sequence 2, Appli
Sequence 31, Appli
Sequence 5704, Ap
Sequence 7, Appli
Sequence 6, Appli
Sequence 12463, A
Sequence 168, App
Sequence 7821, Ap
Sequence 97, Appli
Sequence 7627, Ap
Sequence 110, App
Sequence 102, App
Sequence 38, Appli
Sequence 28, Appli
Sequence 6, Appli
Sequence 16, Appli
Sequence 11489, A

16 76 7.8 332 12 US-09-882-227-440
17 76 7.8 391 12 US-10-057-531A-7
18 76 7.8 391 12 US-10-057-532A-7
19 76 7.8 393 12 US-10-057-531A-3
20 76 7.8 393 12 US-10-057-532A-3
21 76 7.8 394 10 US-09-978-756-3
22 76 7.8 431 12 US-10-057-531A-2
23 76 7.8 431 12 US-10-057-532A-2
24 76 7.8 546 12 US-10-057-531A-1
25 76 7.8 546 12 US-10-057-532A-1
26 76 7.8 581 9 US-09-835-270-2
27 75 7.7 307 9 US-09-815-242-11306
28 75 7.7 394 15 US-10-062-803-2
29 75 7.7 600 11 US-09-893-519A-31
30 74.5 7.6 426 9 US-09-815-242-5704
31 74.5 7.6 506 15 US-10-157-223-7
32 74.5 7.6 560 10 US-09-890-813-6
33 74.5 7.6 705 9 US-09-815-242-12463
34 74.5 7.6 715 12 US-10-247-671-168
35 73.5 7.5 479 12 US-10-032-585-7821
36 73.5 7.5 864 12 US-09-738-630-97
37 73.5 7.5 1128 12 US-10-032-585-7627
38 73 7.5 758 15 US-10-127-032-110
39 73 7.5 892 12 US-10-205-219-102
40 73 7.5 1812 9 US-09-775-938A-38
41 73 7.5 2477 12 US-10-360-053-28
42 72.5 7.4 411 9 US-09-755-325-6
43 72.5 7.4 411 12 US-10-429-997-6
44 72.5 7.4 635 12 US-10-319-315-16
45 72.5 7.4 856 9 US-09-815-242-11489

ALIGNMENTS

RESULT 1
US-10-080-113-3
; Sequence 3, Application US/10080113
; Publication No. US20030186027A1
; GENERAL INFORMATION:
; APPLICANT: SACHS, GEORGE
; APPLICANT: VOLAND, PETRA
; TITLE OF INVENTION: COMPOSITIONS, TEST KITS AND METHODS FOR DETECTING
; TITLE OF INVENTION: HELICOBACTER PYLORI
; FILE REFERENCE: 626 06 PA
; CURRENT APPLICATION NUMBER: US/10/080,113
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 192
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-10-080-113-3

Query Match 100.0%; Score 974; DB 12; Length 192;
Best Local Similarity 100.0%; Pred. No. 5.7e-92;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLEKSFUKSKOLFCLGVLMLQACTCPNTSQNSFLQDVPYVWMLQNRSEYITQGVDSH 50
DB 1 MLEKSFUKSKOLFCLGVLMLQACTCPNTSQNSFLQDVPYVWMLQNRSEYITQGVDSH 60
QY 61 IVDGKTEETEKIATKRAIRVAQNIHVHKLKAYLSKTRIKOKITNEMFIQWTOPIYDS 120
DB 61 IVDGKTEETEKIATKRAIRVAQNIHVHKLKAYLSKTRIKOKITNEMFIQWTOPIYDS 120
QY 121 LMNVDRIGIYNPNNEVFALVFARGFDKDALSEGHLKSLDNQAVSILVAKVEIFKDS 180
DB 121 LMNVDRIGIYNPNNEVFALVFARGFDKDALSEGHLKSLDNQAVSILVAKVEIFKDS 180
QY 181 VNYGDVXVPIAM 192
|||||

Db 181 VNYGDKVPIAM 192

US-09-895-913A-8

Sequence 8, Application US/09895913A

Patent No. US20020160456A1

GENERAL INFORMATION:

APPLICANT: Kieanous, Harold

APPLICANT: Al-Garawi, Amal

APPLICANT: Miller, Charles

APPLICANT: Tomb, Jean Francois

APPLICANT: Omen, Raymond P.

TITLE OF INVENTION: Identification of Polynucleotides

TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the

TITLE OF INVENTION: Genome

FILE REFERENCE: 06132/043002

CURRENT APPLICATION NUMBER: US/09/895,913A

CURRENT FILING DATE: 2001-06-29

PRIOR APPLICATION NUMBER: US 08/881,227

PRIOR FILING DATE: 1997-06-24

NUMBER OF SEQ ID NOS: 368

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 8

LENGTH: 192

TYPE: PRT

ORGANISM: Helicobacter pylori

US-09-895-913A-8

Query Match 99.6%; Score 970; DB 10; Length 192;

Best Local Similarity 99.5%; Pred. No. 1,5e-91;

Matches 191; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLEKSLKQLFLCGLVLMQACTPNTSORSFLODVPYVWMLQNRSEYITQGVDSH 60

Db 1 VLEKSLKQLFLCGLVLMQACTPNTSORSFLODVPYVWMLQNRSEYITQGVDSH 60

QY 61 IVDGKTEIEIKATIRVAQNIHVHKLKAYLSKTNRIKOKITNEMFIQMTQPIYDS 120

Db 61 IVDGKTEIEIKATIRVAQNIHVHKLKAYLSKTNRIKOKITNEMFIQMTQPIYDS 120

QY 121 LAMVDRLGIYINPNNEEVFALVRARGFDKDALSEGLHKLKAYLSKTNRIKOKITNEMFIQMTQPIYDS 180

Db 121 LAMVDRLGIYINPNNEEVFALVRARGFDKDALSEGLHKLKAYLSKTNRIKOKITNEMFIQMTQPIYDS 180

QY 181 VNYGDKVPIAM 192

Db 181 VNYGDKVPIAM 192

RESULT 3

US-09-351-794A-2

Sequence 2, Application US/09351794A

Patent No. US20020042382A1

GENERAL INFORMATION:

APPLICANT: DUFFY, PATRICK E.

APPLICANT: OCKENHOUSE, CHRISTIAN F.

TITLE OF INVENTION: SEQUESTIN

FILE REFERENCE: 38644-175519

CURRENT APPLICATION NUMBER: US/09/351,794A

CURRENT FILING DATE: 1999-07-13

PRIOR APPLICATION NUMBER: 08/559,896

PRIOR FILING DATE: 1995-11-17

NUMBER OF SEQ ID NOS: 7

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 2

LENGTH: 652

TYPE: PRT

ORGANISM: Plasmodium falciparum

US-09-351-794A-2

Query Match 8.4%; Score 82; DB 9; Length 652;

Best Local Similarity 22.4%; Pred. No. 11;

Matches 43; Conservative 20; Mismatches 71; Indels 58; Gaps 7;

QY 47 NRSEYITQGVDSHIVDGKTEIEIKATIRVAQNIHVHKLKAYLSKTNRIKOKITNEMFIQMTQPIYDS 89

Db 29 NTSANTNTSTKESHILDESLETFYRDELKNGKEEIEYFKGNIDKSLDSEFHKILLSE 88

QY 90 -----LKEAYLSKTNRIKOKITNEMFIQMTQPIYDSLMNVDRLGIYINPNNEVFAL 141

Db 89 LNKMDKDELHYEELNREQQKIRNNKQINKYKDEIN-----NNNSDQVDKI 140

QY 142 VR-----ARGFDKDALSEGLHKLKAYLSKTNRIKOKITNEMFIQMTQPIYDS 176

Db 141 HRELEKIEKINKMDKDEIDYREELDKMDRDALYSWYIEDISNKNIKDLKNEKET 200

QY 177 FKDSVNYGDKV 188

Db 201 NKDKKKKDDI 212

RESULT 4

US-10-032-585-7646

Sequence 7646, Application US/10032585

Publication No. US20030180953A1

GENERAL INFORMATION:

APPLICANT: Terry, Roemer D.

APPLICANT: Bo, Jiang

APPLICANT: Charles, Boone

APPLICANT: Howard, Bussey

TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery

FILE REFERENCE: 10182-005-999

CURRENT APPLICATION NUMBER: US/10/032,585

CURRENT FILING DATE: 2001-12-20

NUMBER OF SEQ ID NOS: 8000

SOFTWARE: Patentin version 3.1

SEQ ID NO 7646

LENGTH: 1881

TYPE: PRT

ORGANISM: Candida albicans

FEATURE:

NAME/KEY: MISC FEATURE

LOCATION: (1881)..(1881)

OTHER INFORMATION: X-any amino acid

US-10-032-585-7646

Query Match 8.2%; Score 80; DB 12; Length 1881;

Best Local Similarity 23.8%; Pred. No. 78;

Matches 40; Conservative 29; Mismatches 59; Indels 40; Gaps 8;

QY 29 NTSORSFLOD-----VPYVWMLQNRSEYITQGVDSHIVDGKTEIEIKATIR 81

Db 947 DTEENNELMDKLSASVAYNDLKAKSESEETV-----KAKELETUTSK----- 993

QY 82 VAQNIHVHKLKAYLSKTNRIKOKITNEMFIQMTQPIYDSLMNVDRLGIYINPN 135

Db 994 -IDNLEKELKSEQ-SKNELEGQLQITDSTNEKF-----KELEDELKSIKSNKISSQN 1047

QY 136 EVFALVRARGFDKDALSEGLHKLKAYLSKTNRIKOKITNEMFIQMTQPIYDS 175

Db 1048 SELQKLEKTEKDLOAKDEIDKKAETKSNIDNLNLSSEISSQSKLUKE 1095

RESULT 5

US-10-097-111-295

Sequence 295, Application US/10097111

Publication No. US20030138771A1

GENERAL INFORMATION:

APPLICANT: PELLETIER, JERRY

APPLICANT: GROS, PHILIPPE

APPLICANT: DUBOW, MICHAEL

TITLE OF INVENTION: DNA SEQUENCES FROM S. PNEUMONIAE BACTERIOPHAGE EP1 THAT

FILE REFERENCE: 073406-0603

CURRENT APPLICATION NUMBER: US/10/097,111

; CURRENT FILING DATE: 2002-07-24
 ; PRIOR APPLICATION NUMBER: 09/676,412
 ; PRIOR FILING DATE: 2000-09-29
 ; PRIOR APPLICATION NUMBER: 60/157,218
 ; PRIOR FILING DATE: 1999-09-30
 ; NUMBER OF SEQ ID NOS: 552
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 295
 ; LENGTH: 337
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 US-10-097-111-295

Query Match 8.1%; Score 78.5; DB 12; Length 337;
 Best Local Similarity 23.6%; Pred. No. 10;
 Matches 41; Conservative 26; Mismatches 62; Indels 45; Gaps 9;
 QY 26 TQNTSQRNSFLQDV-----PYWMLQN-----RSEYITQGVDSHIVDGKKEIEIK 72
 DB 81 TCGYTSGLTEFVSNVLRNDGGFYGNQWLKENFGTSSEVVRQGVSPFAFRNGRTEKVE- 139
 QY 73 IATKEATIRVAQNIHK-LKEAYLSKTNRI-----KOKITNEMFIQMTQPIYDSLNNVDR 126
 DB 140 -----HKIIPEDLDKRFPHIPMYERKUTDEL-IEFVDVGYDKLHDCJT 183
 QY 127 LGIYNPNNEVFPALVRARGFDKDALSEGHLKMSLDNQAVSIIVAKVEEI-FKD 179
 DB 184 FVVR-NLKGETVF-----FNRRSVRSKPEHQGEDDPKTEFLYQGVQLVAFRD 229

RESULT 6
 US-09-881-752A-174
 ; Sequence 174, Application US/09881752A
 ; Patent No. US20020115078A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kleantous, Harold
 ; APPLICANT: Al-Garawi, Amal
 ; APPLICANT: Miller, Charles
 ; APPLICANT: Tomb, Jean-Francois
 ; APPLICANT: Cohen, Raymond P.
 ; TITLE OF INVENTION: Identification of Polynucleotides
 ; TITLE OF INVENTION: Encoding No. US20020115078A1e1 Helicobacter Polypeptides in the
 ; FILE REFERENCE: 06132/041002
 ; CURRENT APPLICATION NUMBER: US/09/881,752A
 ; PRIOR FILING DATE: 2001-06-15
 ; PRIOR APPLICATION NUMBER: US 08/833,457
 ; PRIOR FILING DATE: 1997-04-01
 ; NUMBER OF SEQ ID NOS: 370
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 174
 ; LENGTH: 470
 ; TYPE: PRT
 ; ORGANISM: Helicobacter pylori
 US-09-881-752A-174

Query Match 8.1%; Score 78.5; DB 10; Length 470;
 Best Local Similarity 32.9%; Pred. No. 17;
 Matches 28; Conservative 13; Mismatches 29; Indels 15; Gaps 3;
 QY 44 MLQNRSEYITQGVDSHIVDGKKEIEIKATKEATIRVAQNIHVHKLKAYLSKTNRIKQ 103
 DB 111 LLDLSQQVASQGVKPLVVDVGK-----EQGITDSMLVAQNIHVLAQGVDLSKI-----Q 162
 QY 104 KITNEMFIQMTQPIYDSLNNVDRIG 128
 DB 163 QNNNEQ-----LYENIMKVMLLG 180

RESULT 7
 US-10-102-469-20
 ; Sequence 20, Application US/10102469
 ; Publication No. US20030192078A1

; GENERAL INFORMATION:
 ; APPLICANT: Fischhoff, David A.
 ; APPLICANT: Perlak, Frederick J.
 ; TITLE OF INVENTION: SYNTHETIC PLANT GENES AND METHOD FOR PREPARATION
 ; FILE REFERENCE: 38-21 (13553) B
 ; CURRENT APPLICATION NUMBER: US/10/102,469
 ; CURRENT FILING DATE: 2002-09-05
 ; PRIOR APPLICATION NUMBER: 08/434,105
 ; PRIOR FILING DATE: 1995-05-03
 ; PRIOR APPLICATION NUMBER: 07/959,506
 ; PRIOR FILING DATE: 1992-10-09
 ; PRIOR APPLICATION NUMBER: 07/476,661
 ; PRIOR FILING DATE: 1990-02-12
 ; PRIOR APPLICATION NUMBER: 07/315,355
 ; PRIOR FILING DATE: 1989-02-24
 ; NUMBER OF SEQ ID NOS: 43
 ; SOFTWARE: Patent In version 3.1
 ; SEQ ID NO 20
 ; LENGTH: 1189
 ; TYPE: PRT
 ; ORGANISM: artificial sequence
 ; FEATURE:
 ; OTHER INFORMATION: synthetic structural gene encoding an insecticidal protein deriv
 ; OTHER INFORMATION: d from B.t. entomocidus
 US-10-102-469-20

Query Match 8.0%; Score 78; DB 12; Length 1189;
 Best Local Similarity 19.8%; Pred. No. 67;
 Matches 41; Conservative 38; Mismatches 46; Indels 82; Gaps 9;
 QY 32 QRNSFLQDVFWMLQNRSEYITQGVDSHIVDGKKEIEIKATKEATIRVAQNIHVHKLK 91
 DB 2 EENNQNCIPYCNLSNPPEVLDDG-----ERISTGNSSIDISLSLVQFLV 46
 QY 92 -----EAYLSKTNRIKOKITNEMFIQMTQPIYDSLNNVDR 125
 DB 47 SNFVPGGGFLVLGLDFWVGIVGFSQWDAFLVQI-----EQLINERIAEFARNA--ALANLE 100
 QY 126 RLG-----IYI-----NPNNEVFPALVRARGFDKDALSEGHLK----- 158
 DB 101 GLGNFNFIYVEAFKWEEDENNEP-----TRTRVIDRFRIDGLLERDIPSPRISGPEVPL 156
 QY 159 MSLDNQAVSILVAKVEEIFKDSVNYGD 185
 DB 157 LSVYAQAANLHLA-----ILRDSVIFGE 179

RESULT 8
 US-10-156-761-12081
 ; Sequence 12081, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 12081
 ; LENGTH: 427
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-12081

```

Query Match      8.0%; Score 77.5; DB 15; Length 427;
Best Local Similarity 27.3%; Pred.No.18;
Matches 39; Conservative 17; Mismatches 50; Indels 37; Gaps 6;

QY 54 QGVDS-----HIVD-----GKTEBI---EKIAT-----KKAIRVA 83
Db      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
89 RGVDTSKLLISGNAHIITPYNTVDKVTSRFLGKRKIIGTGTGRIGPTVADKINRVGRIQ 148
QY 84 ----QNIVHKLEAYLSKTNRIRIKQKTNEMFICMTQPIYDSLMDVRLGIYNENNEVP 139
Db      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
149 DLYDESILTQVEALDGKKQLTKVFNNRAIEAGQVVEELLTYADRLKPVV---ADTVL 205
QY 140 ALVRARGFDKDALSEGLHKMSLD 162
Db      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
206 VLNKALDDDKVLFEGGGQTLLD 228

RESULT 9
US-09-815-242-11310
; Sequence 11310, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Chisen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Garz, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/131,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/259,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11310
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11310

Query Match      8.0%; Score 77.5; DB 9; Length 856;
Best Local Similarity 25.0%; Pred.No.48;
Matches 41; Conservative 22; Mismatches 60; Indels 41; Gaps 8;

QY 52 ITQGVDSHHVGDGKKTEBEIKATKRAIRVAQNIHV-KLKEAYLSKTNRIKQKITNEMF 110
Db      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
690 LLQVLDEGLHTDSKGV---RVDFKNLTILITSNVASGALLEENSEAE--KQKAIKESL 743
QY 111 IOMTQPIY----DSLAVNDRL-----CIYNPNNEEVSFALVR 143
Db      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
744 ROFFKPFNLKDLBIIISFNALDSHAVINIVGILFENIQKVALERGINIT-LDEAKELIA 802
QY 144 ARGFDK----DALSEGHLHKMSLDNQAVSLVAKVEEIFKDSVNY 183
Db      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
803 EAGEDRFYFGAPLRKALYEVNEVDKLAEILEDKVKE--NDSVAF 844

```

; PRIOR APPLICATION NUMBER: 09/496,005
 ; PRIOR FILING DATE: 2000-02-01
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7
 ; LENGTH: 863
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-163-547-7

Query Match
 Best Local Similarity 8.0%; Score 77.5; DB 15; Length 863;
 Matches 32; Conservative 19; Mismatches 45; Indels 49; Gaps 5;
 QY 2 LKSPKSKQLFLCGLGVMLQACTCPTNSRNF-----LQVPYWMQNRS 49
 DB 571 VKSFMSSKHIF-----ALFNTQORNYKYRFLFLACDSQEDVDSW-----KA 614
 QY 50 EYITGVSSSHIVDGKTEIE-----KIATKRATIRVAQNIHVHKLKEY 94
 DB 615 SLLRAGVYDPKSGVGNKKAENDENGOAENFMDPOLRQVETIRNLVDSYMSIINKCIRDL 674
 QY 95 LSKT-----NPIKQKINEMFIOM 113
 DB 675 IPTINHLMINNVKDFINSELLAQL 699

RESULT 12

US-09-823-187-94
 ; Sequence 94, Application US/09823187
 ; Publication No. US20030096952A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Burgees, Catherine
 ; APPLICANT: Gusev, Vladimir Y
 ; APPLICANT: Liu, Xiaohong
 ; APPLICANT: Majumder, Kumud
 ; APPLICANT: Padigaru, Muralidhar
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Shinkets, Richard A
 ; APPLICANT: Spaderna, Steven K
 ; APPLICANT: Spytek, Kimberly
 ; APPLICANT: Taupier, Raymond J
 ; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
 ; FILE REFERENCE: 15966-745
 ; CURRENT APPLICATION NUMBER: US/09/823,187
 ; PRIOR FILING DATE: 2001-03-29
 ; PRIOR APPLICATION NUMBER: 60/193,339
 ; PRIOR FILING DATE: 2000-03-30
 ; PRIOR APPLICATION NUMBER: 60/193,205
 ; PRIOR FILING DATE: 2000-03-30
 ; PRIOR APPLICATION NUMBER: 60/195,343
 ; PRIOR FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: 60/195,088
 ; PRIOR FILING DATE: 2000-04-06
 ; PRIOR APPLICATION NUMBER: 60/195,005
 ; PRIOR FILING DATE: 2000-04-06
 ; PRIOR APPLICATION NUMBER: 60/195,792
 ; PRIOR FILING DATE: 2000-04-10
 ; PRIOR APPLICATION NUMBER: 60/196,556
 ; PRIOR FILING DATE: 2000-04-11
 ; PRIOR APPLICATION NUMBER: 60/197,081
 ; PRIOR FILING DATE: 2000-04-13
 ; PRIOR APPLICATION NUMBER: 60/197,525
 ; PRIOR FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/197,087
 ; PRIOR FILING DATE: 2000-04-14
 ; NUMBER OF SEQ ID NOS: 103
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 94
 ; LENGTH: 779
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-09-823-187-94

Query Match
 Best Local Similarity 7.9%; Score 76.5; DB 11; Length 779;
 Matches 41; Conservative 29; Mismatches 61; Indels 47; Gaps 9;
 QY 11 QLFCLGLGVMLQACTC-----PN-TSQRNSFLQD---VPYWMQNRSVITQGVSS 59
 DB 204 QLDVMSLGV-VLYVLVCGSLPFGDGNLPTLRQVLEGRFR1PFFWSQDCETLIRML--- 259
 QY 60 HIVDGKTEIEKIATKRATIRVAQNIHVHKLKEYALSKTNRIKQKITNEMFIOMTQPIYD 119
 DB 260 -VVDPAK-----RITIAQIRQHRWQA-----DPTLLQDDPAFD 293
 QY 120 SLMNVDLGLYINPNNEVPALVRASGFKDALSGLHKWSLDNOAV--SILVAVER 175
 DB 294 MQGYTSNLGDIY---NEQVLGIMQALGIDRQRTIESLQNSYNHFAAYLLLELKE 347

RESULT 13

US-09-840-787-41
 ; Sequence 41, Application US/09840787
 ; Patent No. US20020058264A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lal, Preeti
 ; Hillman, Jennifer L.
 ; Bandman, Olga
 ; Shah, Purvi
 ; Au-Young, Janice
 ; Yue, Henry
 ; Guegler, Karl J.
 ; Corley, Neil C.
 ; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
 ; NUMBER OF SEQUENCES: 98
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/840,787
 ; FILING DATE: 23-Apr-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/518,865
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0356 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; TELEX: <Unknown>
 ; INFORMATION FOR SEQ ID NO: 41:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 222 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: SININOT01
 ; CLONE: 2184712
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 41 :
 US-09-840-787-41

Query Match

7.8%; Score 76; DB 9; Length 222;

```

      : : : ||| | : : : | : : : | : : : |
209 GLAIHNNELGLE-----RGLDAALSLAVNGRLVWFSHSIEDRVKLFPMRK 257

RESULT 15
US-10-012-542-141
; Sequence 141, Application US/10012542
; Publication No. US20030044851a1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins

```

```

1 PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509
2 PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
3 PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510
4 PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
5 PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112
6 PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
7 PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113
8 PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
9 NUMBER OF SEQ ID NOS: 532
10 SOFTWARE: PatentIn Ver. 2.0

```

Query March	7.8%; Score 76; DB 15; Length 325;
Best Local Similarity	24.2%; Pred. No. 18;
Matches	56; Conservative 22; Mismatches 57; Indels 96; Gaps
QY	20 LMIQACTCNTSTQRNS-----FLOD-----VP
DB	11 LAIQNC-CFLMWRHSADRCQAVLADERSPOGASQEDIANFQVLKVLPMVWTVLVP
QY	42 YWML--QNSSEYITQGVDSHIVPGKXTEFIEKI-ATKRAITRV---AQNIVHKLKEAYL
DB	70 YWVVFQMGSTVYLGQ---HL-----HIPNIFPANPANISVALRAQGSSTYIPEAWL
QY	96 SKTN-----RHK-----QKTNEMFIQMTQFIYDSLMMYVDRLGI
DB	120 LLANVVVVLIVPLKDRLLIDPLLRCKLLPSAQKALGMFGFTSVI VAGVLEMR.LH-
QY	130 YINPNNEEVALVRARGPDKALSEGHLK---MSLDNQAVSIILVAKVEEIF 177
DB	179 YIH-HNETV-----SIOIGEVLYNAAPSIYMWQIPOVYLLIGISIF 218

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:30:22 ; Search time 12.8168 Seconds
(without alignments)
1440.645 Million cell updates/sec

Title: US-10-080-113-3

Perfect score: 974

Sequence: 1 MLEKSFLLKQLFLCGLGLV.....VEIFKDSVNYGVKVPPIAM 192

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168692 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	974	100.0	192	D64594	hypothetical prote
2	929	95.4	192	E71917	hypothetical prote
3	95	9.8	353	T33782	hypothetical prote
4	94	9.7	315	H31030	H+-transporting tw
5	94	9.7	790	S67803	probable membrane
6	93.5	9.6	294	D90578	elongation factor
7	92.5	9.5	355	T43034	hypothetical prote
8	92.5	9.5	2094	S33124	tpr protein - huma
9	92	9.4	315	A21807	ATP synthase chain
10	90	9.2	6669	S55024	nebulin, skeletal
11	89.5	9.2	586	C90524	hypothetical prote
12	89	9.1	461	T16151	hypothetical prote
13	89	9.1	791	D82901	ATP-dependent prot
14	88.5	9.1	672	E84073	penicillin-binding
15	88.5	9.1	1156	D70356	chromosome assembl
16	88	9.0	293	F49846	sp33 protein - Sh
17	88	9.0	1447	D92909	hypothetical prote
18	87.5	9.0	394	C86630	prophage p61 prote
19	87.5	9.0	1418	D84938	hypothetical prote
20	86.5	8.9	669	C96630	hypothetical prote
21	86.5	8.9	1691	T15564	hypothetical prote
22	86	8.8	1937	T18055	myosin heavy chain
23	85	8.7	422	F75151	hypothetical prote
24	85	8.7	843	S31272	DNA excision repai
25	85	8.7	1244	D873731	probable lipoprote
26	85	8.7	1938	D859233	skeletal myosin he
27	84.5	8.7	527	D33068	myosin heavy chain
28	84.5	8.7	1940	D459287	myosin heavy chain
29	84	8.6	303	B64314	hypothetical prote

30 84 8.6 424 2 T19478
31 83.5 8.6 710 2 A90011
32 83.5 8.6 834 2 B69380
33 83 8.5 332 2 A70228
34 82.5 8.5 518 2 C82502
35 82.5 8.5 643 2 C97174
36 82.5 8.5 715 2 H71489
37 82.5 8.5 900 2 T41607
38 82.5 8.5 1233 2 T30534
39 82.5 8.5 2819 2 A90551
40 82 8.4 230 2 A71603
41 82 8.4 650 2 AB1738
42 82 8.4 837 2 JN0292
43 82 8.4 934 2 S75833
44 82 8.4 958 2 H84783
45 81.5 8.4 243 2 AD1325

hypothetical prote
hypothetical prote
methyl-accepting c
expD protein homol
conserved hypothet
uncharacterized co
probable transcrip
probable vacuolar
chromosome segrega
conserved hypothet
probable integral
probable Na+/H+ an
antigen 332 - mala
hypothetical prote
probable PHD-type
transcription regu

ALIGNMENTS

RESULT 1

D64594

Hypothetical protein HP0596 - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori

C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999

C:Accession: D64594

R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenn
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L
Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: D64594

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-192 <TOM>

A:Cross-references: GB:AE000573; GB:AE000511; NID:92313712; PIDN:AD07665.1; PID:G23137.

C:Genetics:

A:Start codon: GTG

Query Match 100.0%; Score 974; DB 2; Length 192;
Best Local Similarity 100.0%; Pred. No. 5.4e-71;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLEKSFLLKQLFLCGLGLVLMQACTCPNTSQNSFLQDVPYVWMLQNRSEYITQGVDSH 60

Db 1 MLEKSFLLKQLFLCGLGLVLMQACTCPNTSQNSFLQDVPYVWMLQNRSEYITQGVDSH 60

QY 61 IVDGKTEETEEKATKATIRVAQNTVHKLKAYLSKTNRIKQKINEMFIQWTOPIYDS 120

Db 61 IVDGKTEETEEKATKATIRVAQNTVHKLKAYLSKTNRIKQKINEMFIQWTOPIYDS 120

QY 121 LMNVDRIGIYNPNNEVFALVRGDFDKDALSEGLHKMSLDNQAVSILVAKVEIFKDS 180

Db 121 LMNVDRIGIYNPNNEVFALVRGDFDKDALSEGLHKMSLDNQAVSILVAKVEIFKDS 180

QY 181 VNYGDKVKPIAM 192

Db 181 VNYGDKVKPIAM 192

RESULT 2

E71917

Hypothetical protein jhp0543 - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: Strain J99

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C:Accession: E71917

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;

; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
 A:Reference number: A71800; MUID:99120557; PMID:9923682
 A:Accession: E71917
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-192 <ARN>
 A:Cross-references: GB:AE001487; GB:AE001439; NID:g14155086; PIDN:AAD06128.1; PID:g1415508
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: jhp0543

Query Match 95.4%; Score 929; DB 2; Length 192;
 Best Local Similarity 94.3%; Pred. No. 2.2e-67;
 Matches 181; Conservative 8; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MLEKSFKSKQLFLGVLGVLMLQACTCPNTSQRNSFLQDVPYMWLQNRSEYITQGVDSH 60
 DB 1 MLEKSFKSKQLFLGVLGVLMLQACTCPNTSQRNSFLQDVPYMWLQNRSEYITQGVDSH 60
 QY 61 IVGKKTEETKATKATIRVAQNIHVHLKEAYLSKTRIKOKITNEMFIQMTQPIYDS 120
 DB 61 IVGKATEETKATKATIRVAQNIHVHLKEAYLSKTRIKOKITNEMFIQMTQPIYDS 120
 QY 121 LMNVDRGLGIYINPNNEVEFALVRARGFDKDALSEGHLKMSLDNQAVSILVAKVEIFPKDS 180
 DB 121 LMNVDRGLGIYINPNNEVEFALVRARGFDKDALSEGHLKMSLDNQAVSILVAKVEIFPKDS 180
 QY 181 VNYGDKVPIAM 192
 DB 181 VNYGDKVPIAM 192

RESULT 3
 T33782
 hypothetical protein C39F7.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 04-Mar-2000
 C:Accession: T33782
 R:Maggi, L.; Scheet, P.; Dubbalde, C.
 submitted to the EMBL Data Library, October 1998
 A:Description: The sequence of C. elegans cosmid C39F7.
 A:Reference number: Z21407
 A:Accession: T33782
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-353 <MAG>
 A:Cross-references: EMBL:AF101310; PIDN:AAC69214.1; GSPDB:GN00023; CESP:C39F7.5
 A:Experimental source: strain Bristol N2; clone C39F7
 C:Genetics:
 A:Gene: CESP:C39F7.5
 A:Map position: 5
 A:Introns: 14/2; 45/3; 224/1
 C:Superfamily: Caenorhabditis elegans hypothetical protein C39F7.5

Query Match 9.8%; Score 95; DB 2; Length 353;
 Best Local Similarity 25.3%; Pred. No. 3.3;
 Matches 52; Conservative 32; Mismatches 67; Indels 56; Gaps 10;
 QY 1 MLEKSFKSKQ---LFLC-----GLGVLMLQACTCPNTSQRNSFLQDVPYMWLQ 46
 DB 151 LCEQKFLKFPQDAVEIFLKSIIYNEQWPNLGSEMLEQ-LCTDFGCFKKLEEDVKHVKQ 209
 QY 47 NRSEYITQGVDSHIVDGKKTEIEK-----IATKRTIRVAQNIHVHLKEAYL-S 96
 DB 210 D-----LKHGDLIIIVDVGKNTENSESFIRCSSIIAASRSKI-IRSLRLKLEKNEGS 264
 QY 97 KTRIRIKOKITNEM-FIQMTQPIYDSLMNVDRGLGIYINPNNEVEFA----- 140
 DB 265 STSRPRKRIEVEIFLQAFAPFVFLYTRDLWDLAPKSDSISLSQAKAITNSQRT 324
 QY 141 ---LVR-----ARGFDKDALSEGHL 157
 DB 325 DQLARALQLIEARFEVEQLVQGIY 351

RESULT 4

H31090
 H+-transporting two-sector ATPase (EC 3.6.3.14) gamma chain - Anabaena sp. (strain PCC 7
 N:Alternate names: Atpase gamma chain
 C:Species: Anabaena sp.
 A:Variety: PCC 7120
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 03-Jun-2002
 C:Accession: H31090
 R:McCarn, D.F.; Whitaker, R.A.; Alam, J.; Vrba, J.M.; Curtiss, S.E.
 J. Bacteriol. 170, 3448-3458, 1988
 A:Title: Genes encoding the alpha gamma, delta, and four F-0 subunits of ATP synthase c
 A:Reference number: A91875; MUID:88296650; PMID:2500236
 A:Accession: H31090
 A:Molecule type: mRNA
 A:Residues: 1-315 <MCC>
 A:Cross-references: GB:M21659; NID:g141996; PIDN:AAA21992.1; PID:g142004
 A:Experimental source: Anabaena sp. PCC 7120
 C:Genetics:
 A:Gene: atpC
 C:Superfamily: H+-transporting ATP synthase gamma chain
 C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; peripheral membran

Query Match 9.7%; Score 94; DB 2; Length 315;
 Best Local Similarity 22.2%; Pred. No. 3.5;
 Matches 42; Conservative 37; Mismatches 56; Indels 54; Gaps 8;

QY 46 QNRS-EXITQGVDSHIVDGKKTEE-----IEKIATKRTIRVAQNIHVH 89
 DB 102 ENRAKELKAGLDYTFVIVGRKAEQVFRREQPIDASYTGLEQIPTADE---ANKIAD 157
 QY 90 LKEAYLS-KTRIRIKOKITNEMFIQMTQPIYDSLMNVDRGLGIYINPNNEVEFALV----- 142
 DB 158 LLSLFLSEKVDRIELVYTRFVLSVSRPVTQTLPLDTQG--LEAADDEIFLRTTRGGQF 215
 QY 143 -----RARGFDKDALSE-----GLHMSLDNQ-----AVSILVAKVEEIF 177
 DB 216 QVERQTQVTSQARPLPRDSIFEQDPVQLDSLLPLYSNQLLALQESAELAAARTAMS 275
 QY 178 KDSVNYGVDV 186
 DB 276 NASENAGEL 284

RESULT 5

S67803
 Probable membrane protein YDL239c - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein D0771
 C:Species: Saccharomyces cerevisiae
 C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 19-Apr-2002
 C:Accession: S67803
 R:Alt-Moerbe, J.; Schneider, C.; Moro, M.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S67798
 A:Accession: S67803
 A:Molecule type: DNA
 A:Residues: 1-790 <ALT>
 A:Cross-references: EMBL:Z74287; NID:g1431405; PID:e253369; PID:g1431406; GSPDB:GN00004
 A:Experimental source: strain S288C
 C:Genetics:
 A:Gene: SGD:ADY3; MIPS:YDL239C
 A:Cross-references: SGD:S0002398
 A:Map position: 4L
 C:Keywords: transmembrane protein
 F:706-722/Domain: transmembrane #status predicted <TM>

Query Match 9.7%; Score 94; DB 2; Length 790;
 Best Local Similarity 21.8%; Pred. No. 11;
 Matches 52; Conservative 41; Mismatches 79; Indels 66; Gaps 11;

QY 5 SFLKSKQLFLGVLMLQACTCPNTSQ--RNSFLQDVPYMW---LQNRSEYITQGVDS 58

Db 433 SYMQLSLPEKDLGKILBEMKCGHSPFQNGFAKLYPFDQDIXNLENWEGY-KQLKGG 491
Qy 59 SHIVDGGKTEIEIKIATKRATIR-----VAQNIHVHKLKEAYLSKTNRIKQKITNEMFTOM 113
Db 492 IELLEKNDRIRLEKIISVFKLINERLHFQMQQOHSKIK--YLQ-----KEALTKEQQPFL 544
Qy 114 TQIYDLSLMNV-----DLGIGYINPNNEEV-----PAL 141
Db 545 EKRWHDILNKEENFQKLSELKGLILSEKIQKNAEDKLANDYNEHQEIVEKLNQAL 604
Qy 142 VRARGFDKALSGLHKMSLDNOA-----VSLVAKVEEIPKDSVN-----YGD 185
Db 605 IASRWSTQIOESENTHKITDELAKGQSEILKLEETILSLK-EDVFQEKLNKLKLYGD 661

RESULT 6
D90378
elongation factor ts (ef-ts) [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: D90578
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Status: preliminary
A:Accession: D90578
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-294 <KUR>
A:Cross-references: GB:AL445566; PID:g14089947; PIDN:CAC13705.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPU 5320
A:Genetic code: SOC3
C:Superfamily: translation elongation factor EF-Ts

Query Match 9.6%; Score 93.5; DB 2; Length 294;
Best Local Similarity 21.4%; Pred. No. 3.5;
Matches 41; Conservative 43; Mismatches 69; Indels 39; Gaps 9;

Qy 1 MLEKSF--LKSQKFLCGLGVLMLQACT--CPNFSQNSRFLQDVPFWM/LONS-----EVI 52
Db 101 LVENSFQSMESAENIFENDLTILEATTKATATIGEKISFRRAKFDLLEDQTIQAYTHA 160
Qy 53 TQGVDSHIVDGGKTEIEIKIATKRATIR-----RVAQNIHVHKLKEAYL-----SKT 98
Db 161 NGRASFLYRGKNEEVAKVAMHIAAMNPEYMSANEVQEKIEKLAFLKSPALAGKP 220
Qy 99 NRIKQKITNEM-----FIQMTOP-IYDSLMNVDLGIGYINPNNEEVFALVRARGFDKD 150
Db 221 EKIQCSILSGMLNKAALAEFVLLNQPFVMSLSVEQ---YLKNKSEALEMIRYE----- 272
Qy 151 ALSBGLHKMSLD 162
Db 273 -VGEIGIEKKAVD 283

RESULT 7
T43094
hypothetical protein - Lactococcus lactis plasmid pMRC01
C:Species: Lactococcus lactis
C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
C:Accession: T43094
R:Dougherty, B.A.; Hill, C.; Weidman, J.F.; Richardson, D.R.; Venter, J.C.; Ross, R.P.
Mol. Microbiol. 29, 1029-1038, 1998
A:Title: Sequence and analysis of the 60 kb conjugative, bacteriocin-producing plasmid p
A:Reference number: 222314
A:Accession: T43094
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-355 <DOU>
A:Cross-references: EMBL:AE001272; PIDN:AAC56005.1
A:Experimental source: strain DFC147

C:Genetics:
A:Genome: plasmid pMRC01
A:Note: ORF00022
C:Superfamily: Lactococcus lactis hypothetical protein PL08711.5

Query Match 9.5%; Score 92.5; DB 2; Length 355;
Best Local Similarity 24.5%; Pred. No. 5.3;
Matches 38; Conservative 28; Mismatches 50; Indels 39; Gaps 6;

Qy 43 WMLONRSEVITQGVDSHIVDGGK--TEIEIKIATKRATIRVAQNIHVHKLKEAYLSKTNR 100
Db 44 YSINNINLYSQNPEDATHVAGFQMGTDNFKVKNKGEKAIKRAIAPIIKK-----LSESEK 98
Qy 101 IKKQITNEMFIQMTQ--PIYD-----SLMNV-----RLGIYINP 133
Db 99 IKLKTDERAIVGYRYPVDFVSQTSQDPLPSARDFVKENLSEVENVDVLYKSLKNYINQ 158

Qy 134 N-----NEEVFALVRARGFDKALSGLHKMSLDN 163
Db 159 NTDIKVSEELSDFEVKGFFRPSTNQINMESVDN 193

RESULT 8
S33124
tpr protein - human
N:Alternate names: kinase-related transforming protein (tpr-met); protein with promoter
C:Species: Homo sapiens (man)
C>Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 11-Jan-2002
C:Accession: S33124; S23740; S00928; G01185
R:Mitchell, P.J.; Cooper, C.S.
Oncogene 7, 2329-2333, 1992
A:Title: The human tpr gene encodes a protein of 2094 amino acids that has extensive con
A:Reference number: S33124; MUID:93064711; PMID:1437155
A:Accession: S33124
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-2094 <MIT>
A:Cross-references: EMBL:X66397; NID:G633225
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991
R:Mitchell, P.J.; Cooper, C.S.
Oncogene 7, 383-388, 1992
A:Title: Nucleotide sequence analysis of human tpr cDNA clones.
A:Reference number: S23740; MUID:92195670; PMID:1549355
A:Accession: S23740
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-725, 'L' <MIT>
A:Cross-references: EMBL:X63105; NID:G37257; PIDN:CAA44819.1; PID:G37258
R:King, H.W.S.; Tempest, P.R.; Merrifield, K.R.; Rance, A.J.
Oncogene 2, 617-619, 1988
A:Title: Tpr homologues activate met and raf.
A:Reference number: S00928; MUID:88262257; PMID:3387099
A:Accession: S00928
A:Molecule type: mRNA
A:Residues: 1-31, 'R', 33-142 <KIN>
A:Cross-references: EMBL:Y00672; NID:G37255; PIDN:CAA68681.1; PID:G37256
R:Greco, A.
submitted to the EMBL Data Library, December 1995
A:Reference number: H00592
A:Accession: G01185
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 144-228 <GR>
A:Cross-references: EMBL:X94208; NID:g1296797; PIDN:CAA63904.1; PID:g1296798
C:Genetics:
A:Gene: GDB:TPR
A:Cross-references: GDB:128821; OMIM:189940
A:Map position: 1q25-1q25
A:Introns: 177/3

Query Match 9.5%; Score 92.5; DB 2; Length 2094;
Best Local Similarity 24.1%; Pred. No. 47;
Matches 48; Conservative 38; Mismatches 39; Indels 10;

QY 17 LGVLMQACTCPTNRSNFIQ-----DVPYMLQNRSEYITQGV-----S 58
 Db 1302 LDLPLOEANA-ELSEKGMQAEKKLEEDVQKWKARNQ-HLVSQQKDPDTSEYRKLLS 1359
 QY 59 SHIVGKK-----TEI-----EKIATKRAIRVAQNIHVKLKEAYLSKTNRIKOKITNEMF 110
 Db 1360 EREVHTKIQQTETIGRLKAEIARSNASLTNNQNLQSLKED-LNKVTEKETIQKLDL 1418
 QY 111 IQM--TQPIYDLMVDRGLGIYINPNEEVPALVRARGFDKALSEGHLHMSLDNQAISI 168
 Db 1419 AKIIDQKVKITQVKIGRRYKTYEELKA-----QODKWMESAQSSGDHQEHV 1471
 QY 169 LVAKVEEFTKSVNYGVK 187
 Db 1472 SVQEMQEL-KETLNOAETK 1489
 RESULT 9
 AD1807
 ATP synthase chain gamma [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AD1807
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete genomic sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120
 A:Reference number: AB1807; MUID:21595285; PMID:111759840
 A:Accession: AD1807
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-315 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA077528.1; PID:g17134982; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: atpC
 C:Superfamily: H+-transporting ATP synthase gamma chain
 Query Match 9.4%; Score 92; DB 2; Length 315;
 Best Local Similarity 21.7%; Pred. No. 5;
 Matches 41; Conservative 37; Mismatches 57; Indels 54; Gaps 8;
 QY 46 QNRS-EYITQGVDSHIVDGKTEE-----LEKIATKRAIRVAQNIHVK 89
 Db 102 ENRAKELKAEGLDYTFVIGRAEQYFRREQPIDASYTGLEQIPTADE-----ANKIAD 157
 QY 90 LKEAYLS-KTNRIKOKITNEMFIQNTQPIYDSLMMVDRGLGIYINPNEEVPALVRARG-- 146
 Db 158 LLSLFSEKVDRIELVYFVSVLSVSRPVIQTLLPLDTQG--LEAADDEIFRLTRGGQF 215
 QY 147 -----FKDALS--EGLKMSLDNO-----AVSILVAKVEEFT 177
 Db 216 QVETVTSQARPLPRDMIFQDPVQILDLSLPLYSNQLRALQESAASELAARTAMS 275
 QY 178 KDSVNYGV 186
 Db 276 NASENAGEL 284
 RESULT 10
 S55024
 nebulin, skeletal muscle - human
 C:Species: Homo sapiens (man)
 C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 05-Nov-1999
 C:Accession: S55024; S55025; A40847
 R:Labelit, S.
 submitted to the EMBL Data Library, January 1995
 A:Reference number: S55024
 A:Accession: S55024
 A:Molecule type: mRNA
 A:Residues: 1-6669 <LAB1>

A:Cross-references: EMBL:X83957; NID:g806561; PIDN:CAA58788.1; PID:g806562
 R:Labelit, S.; Kolmerer, B.
 J. Mol. Biol. 248, 308-315, 1995
 A:Title: The complete primary structure of human nebulin and its correlation to muscle S
 A:Reference number: S55025; MUID:95257391; PMID:7739042
 A:Accession: S55025
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 3731-3930; 6501-6669 <LAB2>
 A:Cross-references: EMBL:X83957
 R:Jin, J.P.; Wang, K.
 J. Biol. Chem. 265, 21215-21223, 1991
 A:Title: Cloning, expression, and protein interaction of human nebulin fragments compose
 A:Reference number: A40847; MUID:92042000; PMID:1882316
 A:Accession: A40847
 A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
 A:Molecule type: mRNA
 A:Residues: 6263-6271, 'M', 6273-6329 <JIN>
 C:Genetics:
 A:Gene: GDB:NEB
 A:Cross-references: GDB:120224; OMIM:161650
 A:Map position: 2824.1-2824.2
 C:Superfamily: human skeletal muscle nebulin; SH3 homology
 C:Keywords: alternative splicing; skeletal muscle; tandem repeat
 F:78-6531/Region: 35-residue repeats
 F:6617-6666/Domain: SH3 homology <SH3>
 Query Match 9.2%; Score 90; DB 2; Length 6669;
 Best Local Similarity 20.0%; Pred. No. 3; le-02;
 Matches 45; Conservative 40; Mismatches 80; Indels 60; Gaps 9;
 QY 3 EKSLKSKQLP---LCGLGVLMQACTCPTNRSNFIQNTQPIYDSLMMVDRGLGIYINPNEEVPALVRARGFKD 56
 Db 2833 KQPEKWKTKFSGPDMGLGVLAKEC-----QTLVSDVDY-----KNYLHQWTCIP 2878
 QY 57 DSSHIVDGKTEETIEKIAKRAIRVAQNI-----VHKLKEA----- 93
 Db 2879 QDSVHARQAYDLSQSNMYSKDLQWGRGIGWISIGSLDVEKCKRATIELSKIRQPPD 2938
 QY 94 ---YLSKTNRIKOKITNEMFIQNTQPIYDSLMMVDRGLGIYINPNEEVPALVRARGFKD 150
 Db 2939 RFKFTSVTDSLEQVLAKNANLNKRLYTEAMKDKTQHIIMPDTPEIM-LARQ---NKI 2994
 QY 151 ALSGLKMS-----LDNQAVSILVAKVEEFTKDSVNYGD 185
 Db 2995 NYSETLKLANEAKKGYDLRSDAIPVAAKASRDVISDYKYKD 3039
 RESULT 11
 C90524
 hypothetical protein MYP 0990 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
 C:Species: Mycoplasma pulmonis
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
 C:Accession: C90524
 R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.
 Nucleic Acids Res. 29, 2145-2153, 2001
 A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pul
 A:Reference number: A99512; MUID:21267165; PMID:11353084
 A:Accession: C90524
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-586 <KUR>
 A:Cross-references: GB:AL445566; PID:g14089512; PIDN:CAC13272.1; GSPDB:GN00153
 A:Experimental source: strain UAB CTIP
 C:Genetics:
 A:Gene: MYP 0990
 A:Genetic code: SGC3
 Query Match 9.2%; Score 89.5; DB 2; Length 586;
 Best Local Similarity 24.1%; Pred. No. 17;
 Matches 40; Conservative 23; Mismatches 62; Indels 41; Gaps 7;
 QY 31 SQNRSFLQDVPYV-----MLQNRSEYITQGVDSHIVDGKTEETIEKIAKRAIRVAQNI 86

Db 135 ANKNVIGDQYQWQKLNFLKNEINEINWKTINLID-KHKKELEKNKVR-----KKL 187
 Qy 87 VHKLEAVLSKTNRIKQKITN-----EMFIQMTQPIYDSLMMVDRGLGIYINPNNEEV 138
 Db 188 IKKYIE--ISKQNSKDKANLLAKITFLSKIFPKDIL-----KQFESL 231
 Qy 139 FALVRAAGFDKDALSEGHLKMSLD-----NOAVSILVAKVEEIFK 178
 Db 232 KDLVESLKQKKLFKSNIKFKPKQDKEDLNKKNKIKKIINKKVEEAAK 277

RESULT 12
 T16161
 hypothetical protein F26A1.2 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T16161
 R:Fulton, L.
 submitted to the EMBL Data Library, May 1995
 A:Description: The sequence of *C. elegans* cosmid F26A1.
 A:Reference number: Z18469
 A:Accession: T16161
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-461 <FUL>
 A:Cross-references: EMBL:U27312; NID:G860679; PID:G860690; PIDN:AAA68254.1; CESP:F26A1.2
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:F26A1.2
 A:Introns: 203/2; 274/1

Query Match 9.1%; Score 89; DB 2; Length 461;
 Best Local Similarity 24.2%; Pred. No. 14;
 Matches 30; Conservative 23; Mismatches 47; Indels 24; Gaps 4;

Qy 84 QNIVHKLKAYLSKTNRIKQKITNEMFTQMTQPIYDSLMM-----NVDRLGIYINPNNEEV 139
 Db 318 KRIIPLETRVFLFHWQEHKATBPMFPPEVPVYSTSLSCVENISSYGGVGNDGLY 377

Qy 140 AL-----VRAGFDKDALSEGHLKMSLDNOAVSILVAKVE-----EIFKDSVNY 183
 Db 378 DLSDRQIICFHKNIRA-----YHSLQKCHKNYSNNWTPSLIYVOEAVMBKEIVEDDIRN 433

Qy 184 GDNK 187
 Db 434 GTIE 437

RESULT 13
 D82901
 ATP-dependent proteinase U1348 [imported] - *Ureaplasma urealyticum*
 C:Species: *Ureaplasma urealyticum*
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: D82901
 R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
 submitted to GenBank, February 2000
 A:Description: The complete sequence of *Ureaplasma urealyticum*: Alternate views of a min
 A:Reference number: A82870
 A:Accession: D82901
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-791 <GLA>
 A:Cross-references: GB:AE002132; GB:AF222894; NID:G6899327; PIDN:AAF30757.1; GSPDB:GN001
 A:Experimental source: serovar 3; biovar 1
 C:Genetics:
 A:Gene: lon; U1348
 A:Genetic code: SGC3
 C:Superfamily: ATP-dependent serine proteinase La

Query Match 9.1%; Score 89; DB 2; Length 791;
 Best Local Similarity 24.1%; Pred. No. 27;
 Matches 35; Conservative 25; Mismatches 59; Indels 26; Gaps 6;

Qy 56 VDSHSHVDGKTEIEKIATKRATIRVAQNI-----VHKLEAYLSKTNRIKQKITNE 108
 Db 51 IDTEVNV---FDLXYVGLVLKIKSIDNFDGYSIEVEGLKAVVINNENDVIDALEYE 107
 Qy 109 MFIQMTQPIYDSLMMVDRGLGIYINPNNEEVFALVRAAGFDKDALSEGHLKMSLDNOAVS 167
 Db 108 YEDVITNPI---LSNKE--IAINGINSEIFNIINRSRHRKNINPDMHALISLEKEKFA 162
 Qy 168 ILVAKV-----BEIFKDSVN 182
 Db 163 YLAAATVINDYDNEISEKTIEDRIN 187

RESULT 14
 E84073
 penicillin-binding protein 3 pbpc [imported] - *Bacillus halodurans* (strain C-125)
 C:Species: *Bacillus halodurans*
 C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: E84073
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hir
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A:Reference number: A83650; NUID:20512582; PMID:11058132
 A:Accession: E84073
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-672 <STO>
 A:Cross-references: GB:AP001518; GB:BA000004; NID:G10175792; PIDN:BA07108.1; GSPDB:GN0
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: pbpc
 C:Superfamily: penicillin-binding protein 2B

Query Match 9.1%; Score 88.5; DB 2; Length 672;
 Best Local Similarity 28.6%; Pred. No. 24;
 Matches 32; Conservative 17; Mismatches 36; Indels 27; Gaps 5;

Qy 77 RATIRVAQNIHVHKLEAYLSKTNRIKQKITNEMFIQMT-----QPIYDSLMMVDRGLY 130
 Db 317 RIFIRTAGD---ELKETLE-----KEAVDGET-VQITIDADVQRSIYDSLADGGGLGWS 367

Qy 131 INPNNEVFALVRAAGFDKDALSEGHLKMSLDNOAVSILVAKVEEIFKDSVN 182
 Db 368 LHPTSGEVLALVSPSYDPNEF-----VVGITSTRYELQEDERN 407

RESULT 15
 B70356
 chromosome assembly protein homolog - *Aquifex aeolicus*
 C:Species: *Aquifex aeolicus*
 C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 02-Jun-2000
 C:Accession: B70356
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O
 V.
 Nature 392, 353-358, 1998
 A:Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.
 A:Reference number: A70300; NUID:98196666; PMID:9537320
 A:Accession: B70356
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1156 <AQF>
 A:Cross-references: GB:AP000699; NID:G2983238; PIDN:AAC06839.1; PID:G2983243; GB:AE0006
 A:Experimental source: strain VF5
 C:Genetics:
 A:Gene: xcpC
 C:Superfamily: chromosome segregation protein SMCI

Query Match 9.1%; Score 88.5; DB 2; Length 1156;
 Best Local Similarity 24.2%; Pred. No. 47;
 Matches 43; Conservative 38; Mismatches 70; Indels 27; Gaps 8;

Qy 30 TSQRNSFLQDPVYVWMLQNRSEY-----ITQGVDSHIVDGKKTETIE----KIATK 76

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:29:11 ; Search time 7.53927 Seconds
(without alignments)
1197.613 Million cell updates/s

Title: US-10-080-113-3

Perfect score: 974
Sequence: 1 MLEKSFLEKSKQLFLCGLGLV.....VEEIFKDSVNYGCDVKVPIAM 192

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt 41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	93.5	9.6	294	1	EFTS_MYCPU	Q98g37 mycoplasma
2	92.5	9.5	249	1	TFR_HUMAN	P12270 homo sapien
3	92	9.4	293	1	SPAO_SHIPL	P35534 shigella fl
4	92	9.4	315	1	ATXP_ANASP	P12408 anabaena sp
5	91.9	9.3	404	1	EXTL_FUSNN	Q8ren3 fusbacteri
6	90.9	9.2	6669	1	NEBU_HUMAN	P20929 homo sapien
7	89	9.1	461	1	YIL2_CAREL	Q19802 casnozhabdi
8	87.5	9.0	1418	1	SM4_YEAST	Q12267 saccharomyc
9	86	8.8	1337	1	MW8_HUMAN	P13535 homo sapien
10	85.5	8.8	219	1	EFTS_GUTH	Q78481 guillardi
11	85	8.7	843	1	RA25_YEAST	Q00578 saccharomyc
12	85	8.7	1244	1	Y307_MYCPN	P75342 mycoplasma
13	85	8.7	1938	1	MVH4_RABIT	Q28641 cryptocolagus
14	84	8.6	303	1	Y114_MBTJA	Q57578 methanococcus
15	82.5	8.5	715	1	GREX_CHLTR	Q84641 chlamydia t
16	82.5	8.5	1261	1	APAF_BRARE	Q949h8 brachydanio
17	81	8.3	421	1	YGUO_YEAST	P53150 saccharomyc
18	81	8.3	709	1	RRP2_INCUJ	P13878 influenza c
19	80.5	8.3	328	1	COAD_THEAC	Q9hiy2 thermoplasm
20	80.5	8.3	771	1	R1R1_VACCV	P20503 vaccinia vi
21	80.5	8.3	771	1	R1R1_VACCV	P12848 vaccinia vi
22	80	8.2	1939	1	MVH4_HUMAN	Q9y623 homo sapien
23	79.5	8.2	318	1	YIL7_YEAST	P40476 saccharomyc
24	79.5	8.2	686	1	GRB_STRCO	P35886 streptomyc
25	79	8.1	414	1	YG35_MBTJA	Q59029 methanococc
26	79	8.1	811	1	R1R1_SCHEP	P36602 schizosacch
27	79	8.1	1557	1	DVAL1_DICVI	Q24702 dictyocaulu
28	79	8.1	2452	1	RPB1_PLAFD	P14248 plasmodium
29	78.5	8.1	457	1	V162_UREPA	Q9p9y2 ureaplasma
30	78	8.0	144	1	V991_MBTJA	Q58398 methanococc
31	78	8.0	503	1	NUSA_RICPR	Q9scz7 rickettsia
32	78	8.0	1197	1	DFOM_PODAN	Q01529 podospora a
33	77.5	8.0	335	1	FEN_THEYO	Q97b98 thermoplasm

Q8mrj6 geodia cydo
P71494 helicobac
P47549 mycoplasma
Q01637 drosophila
O05957 rickettsia
P25457 schizosacch
Q9ZJ29 helicobac
P08450 synechococc
Q8K354 buchnera ap
Q07384 african swi
P57398 buchnera ap
P17634 xenopus lae

ALIGNMENTS

RESULT 1

```

EFTS_MYCPU          PRT;      294 AA.
ID AC Q98Q37;
16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
DE Elongation factor Ts (EF-Ts).
TSF OR MYPU 5320.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN [1]
SEQUENCE FROM N.A.
STRAIN=UAB CTIP;
MEDLINE=21267165; PubMed=11353084;
RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galissou F.,
RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
RE Mycoplasma pulmonis."
RL Nucleic Acids Res. 29:2145-2153(2001).
CC -!- FUNCTION: Associates with the EF-Tu.GDP complex and induces the
CC exchange of GDP to GTP. It remains bound to the aminoacyl-tRNA.EF-
CC Tu.GTP complex up to the GTP hydrolysis stage on the ribosome.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE EF-TS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----

```

CC	EMBL; AL445565; CAC13705.1; -
CC	PIR; D90578; D90578.
CDR	Mypullist; MYPUL 5320; -
CDR	HAMAP; MF 00050; - 1.
CDR	InterPro; IPR001816; EF_TS.
CDR	InterPro; IPR000443; UBA_domain
CDR	InterPro; PFO0689; EF_TS; 1.
CDR	Pfam; PFO0627; UBA; 1.
CDR	Pfam; PFO0627; UBA; 1.
CDR	TIGRFAMS; TIGR00116; tsf; 1.
CDR	PROSITE; PS01126; EF_TS_1; 1.
CDR	PROSITE; PS01127; EF_TS_2; 1.
KMW	Elongation factor; Protein bio
TT	SITE 81
TT	SITE 84
TT	I T
TSO	SEQUENCE 294 AA; 32627 MW;

Query Match	9.6%	Score 93.5;	DB 1;	Length 294;
Best Local Similarity	21.4%	Pred. No. 1.8;		
Matches 41;	Conservative 43;	Mismatches 69;	Indels 3	

9

QY 1 MLEKSF--LKSQFLCGLGVLMLQACT--CPNTSQRSFLQDVPMVWMLQNRS-----EYI 52
 DB 101 LVENSFQMSAEINIFMENDLTILEATTATATIGEXISFRRAKFFDLLEDQIGAYTHA 160
 QY 53 TOGVDSHVVDKTEIEKATKATI-----RVAQNIHKLKEAYL-----SKT 98
 DB 161 NGRASFLVGRKNEEVAKNVAMHIAWNPYMSANEVPEKIEKLAEBFLKSPALAGP 220
 QY 99 NRIKQKINEM-----FIQTOP-IYDSLNVDRGLIYNPNNEVPALVARGFDKD 150
 DB 221 EKIQQSILSGMLNKALEFVLLNQPFVWSSLSVEQ---YLKKNKSEALEMIRVE----- 272
 QY 151 ALSGELHQLMSLD 162
 DB 273 -VGEIGIEKAVD 283

RESULT 2
 ID TPR HUMAN STANDARD; PRT; 2349 AA.
 AC P12270;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nucleoprotein TPR.
 GN TPR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=93064711; PubMed=1437155;
 RX Mitchell P.J., Cooper C.S.;
 RA "The human tpr gene encodes a protein of 2094 amino acids that has
 RT extensive coiled-coil regions and an acidic C-terminal domain."
 RL Oncogene 7:2329-2333(1992).
 RN [2]
 RP REVISIONS, AND CHARACTERIZATION.
 RX MEDLINE=95096166; PubMed=7798308;
 RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
 RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
 RT "Tpr, a large coiled coil protein whose amino terminus is involved in
 RT activation of oncogenic kinases, is localized to the cytoplasmic
 RT surface of the nuclear pore complex."
 RL J. Cell Biol. 127:1515-1526(1994).
 RN [3]
 RP SEQUENCE OF 1-142 FROM N.A.
 RX MEDLINE=88262257; PubMed=3387099;
 RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;
 RT "tpr homologues activate met and raf."
 RL Oncogene 2:617-619(1988).
 CC -!- FUNCTION: COMPONENT OF THE CYTOPLASMIC FIBRILS OF THE NUCLEAR PORE
 CC COMPLEX IMPLICATED IN NUCLEAR PROTEIN IMPORT. ITS AMINO TERMINUS
 CC IS INVOLVED IN ACTIVATION OF ONCOGENIC KINASES.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE
 CC COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH
 CC TRP-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER
 CC COMPONENTS, INCLUDING P62.
 CC -!- TISSUE SPECIFICITY: HIGHEST IN TESTIS, LUNG, THYMUS, SPLEEN AND
 CC BRAIN, LOWER LEVELS IN HEART, LIVER AND KIDNEY.
 CC -!- DISEASE: INVOLVED IN TUMORIGENIC REARRANGEMENTS WITH THE MET, TRK
 CC OR RAF GENES.
 CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
 CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/TPRID282.html".
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC EMBL; X66397; CAA47021.1; -
 DR EMBL; Y00672; CAA69881.1; -
 DR Genew; HGNC:12017; TPR.
 DR MIM; 189940; -
 DR GO; GO:0005737; C:cytoplasm; TAS.
 DR GO; GO:0005643; C:nuclear pore; TAS.
 DR GO; GO:0006605; P:protein-nucleus import; TAS.
 DR GO; GO:0006605; P:protein-nucleus import; TAS.
 KW Coiled coil; Proto-oncogene; Chromosomal translocation;
 KW Nuclear protein; Transport.
 FT DOMAIN 78 360
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 422 571
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 575 628
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 758 805
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 834 869
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 934 979
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1004 1064
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1138 1166
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1196 1241
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1262 1304
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1354 1434
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1476 1595
 FT POLY-SER.
 FT DOMAIN 527 530
 FT POLY-GLU.
 FT DOMAIN 1833 1836
 FT POLY-ASP.
 FT DOMAIN 1957 1964
 FT POLY-SER.
 FT DOMAIN 2295 2298
 FT POLY-SER.
 SQ SEQUENCE 2349 AA; 265600 MW; AFDD6885CEDCA9EF CRC64;
 Query Match 9.5%; Score 92.5; DB 1; Length 2349;
 Best Local Similarity 24.1%; Pred. No. 25;
 Matches 48; Conservative 38; Mismatches 74; Indels 39; Gaps 10;
 QY 17 LGVLMQLQACTCPTSQRSFLQ-----DVPYMLQNRSEYIQGVN-----S 58
 DB 1302 LDILPLQEAANA-ELSEKSGMLQAEKKLLEEDVWRKARNQ-HLVSQQKDPDTEYRKLLS 1359
 QY 59 SHIVDCK-----TEI-----EKIATKATIRVAQNIHKLKEAYLSTKTRIKOKITNEMF 110
 DB 1360 EREVTKRIQQLTEIEGRKAEIARSNASLTNNQLIOSLKD-LNKVTEKETIQKOLD 1418
 QY 111 IQM--TQPIYDSLNVDRGLIYNPNNEVPALVARGFDKDSEGLHKLMSLDNQAVSI 168
 DB 1419 AKIIDQEKVKITQVKKIGRYKTYEELXA-----QDKVMETSAQSGDHEQHV 1471
 QY 169 LVAKVEEIPKDSVNYGDVK 187
 DB 1472 SVQEMOEL-KETLNQAEIK 1489

RESULT 3
 ID SPAO SHIFL STANDARD; PRT; 293 AA.
 AC P35534; Q8VSG8;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Surface presentation of antigens protein spaO (Spa33 protein).
 GN SPAO OR SPA33 OR CPO152.
 OS Shigella flexneri, and
 OS Shigella sonnei.
 OG Plasmid pWR100, Plasmid pMYSH6000, and Plasmid pCP301.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Shigella.
 OX NCBI_TaxID=623, 624;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.flexneri; STRAIN=M90T / Serotype 5; PLASMID=pWR100;
 RX MEDLINE=92193289; PubMed=1312536;
 RA Venkatesan M.M., Buysse J.M., Oaks E.V.;
 RT "Surface presentation of Shigella flexneri invasion plasmid antigens
 RT requires the products of the spa locus."
 RL J. Bacteriol. 174:1990-2001(1992).
 RN [2]

SEQUENCE FROM N.A.
SPECIES=S.flexneri; STRAIN=M90T / Serotype 5; PLASMID=pWR100;
MEDLINE=2056792; PubMed=1115111;
RA Buchrieser C., Glaeser P., Rusnick C., Nedjari H., d'Hauteville H.,
Kunsth F., Sansonetti P., Parsot C.;
RA "The virulence plasmid pWR100 and the repertoire of proteins secreted
RT by the type III secretion apparatus of Shigella flexneri.";
RL Mol. Microbiol. 38:760-771(2000).
[3]
SEQUENCE FROM N.A.
SPECIES=S.flexneri; STRAIN=M90T / Serotype 5; PLASMID=pWR100;
RX MEDLINE=21189246; PubMed=11292750;
RA Venkatesan M.M., Goldberg M.B., Rose D.J., Grotbeck E.J., Burland V.,
RA Blatner F.R.;
RT "Complete DNA sequence and analysis of the large virulence plasmid of
RT Shigella flexneri";
RL Infect. Immun. 69:3271-3285(2001).
[4]
SEQUENCE FROM N.A.
SPECIES=S.flexneri; STRAIN=YSH6000 / Serotype 2a; PLASMID=pYSH6000;
RX MEDLINE=93224456; PubMed=8385666;
RA Sasakawa C., Komatsu K., Iobe T., Suzuki T., Yoshikawa M.;
RA "Eight genes in region 5 that form an operon are essential for
RT invasion of epithelial cells by Shigella flexneri 2a";
RL J. Bacteriol. 175:2334-2346(1993).
[5]
SEQUENCE FROM N.A.
SPECIES=S.flexneri; STRAIN=301 / Serotype 2a; PLASMID=pCF301;
RX MEDLINE=2272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157";
RL Nucleic Acids Res. 30:4432-4441(2002).
[6]
SEQUENCE FROM N.A.
SPECIES=S.sonnei; STRAIN=HW383;
RA Arakawa E., Kato J.I., Ito K.I., Watanabe H.;
RT "Comparison and high conservation of nucleotide sequences of spa-mxi
RT regions between S.sonnei and S.flexneri -- identification of a new
RT gene coding plausible membrane protein";
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -! FUNCTION: REQUIRED FOR SURFACE PRESENTATION OF INVASION PLASMID
CC ANTIGENS. COULD PLAY A ROLE IN PRESERVING THE TRANSLOCATION
CC COMPETENCE OF THE IPA ANTIGENS. REQUIRED FOR INVASION AND FOR
CC SECRETION OF THE THREE IPA PROTEINS.
CC -! SIMILARITY: BELONGS TO THE FLIN/MOPA/SPAO FAMILY.
CC -! CAUTION: Ref.5 sequence differs from that shown due to a stop
CC codon in position 12.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC between the Swiss Institute of Bioinformatics and the EMBL outstat-
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commer-
CC entities requires a license agreement (See <http://www.isb-sib.ch/annou->
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81458; AAA26543.1; --
DR EMBL; AL391753; CAC05827.1; --
DR EMBL; AF348706; AAK18471.1; --
DR EMBL; D13663; BAA02828.1; --
DR EMBL; AF386526; AAL72304.1; ALT_SEQ.
DR EMBL; D50601; BAA09161.1; --
DR FIR; F49846; F49846.
DR InterPro; IPR003283; SecYIIMPO.
DR InterPro; IPR001543; Spoa.
DR Pfam; PF01052; Spoa; 1.
DR PRINTS; PR01339; TYPE3OMOPROT.
DR ProDom; PD001777; Spoa; 1.

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF242564; AAA21992.1; -
EMBL; AP003581; BAB7528.1; -
PIR; AD1807; AD1807.
InterPro: IPR000131; ATPase_gamma.
PRINTS; PR00126; ATPase_gamma.
TIGRFAMs; TIGR01146; ATPsyn_Figamma; 1.
PROSITE; PS00153; ATPase_gamma; 1.
ATP synthesis; CF(1); Hydrogen ion transport; Hydrolase;
Complete proteome.
CONFLICT 233 M -> S (IN REF. 1).
SEQUENCE 315 AA; 35277 MW; AA895589D612424 CRC64;

Query Match 9.4%; Score 92; DB 1; Length 315;
Best Local Similarity 21.7%; Pred. No. 2.5;
Matches 41; Conservative 37; Mismatches 57; Indels 54; Gaps 8;

QY 46 QNRG-EVITQGVDSHIVDGKTEE-----IEKIATKRAIRVAQIVHVK 89
DB 102 ENRAKELKAEGLDTFVIVGKAEQYFRREQPIDASYTGLEQIPADE-----ANKIADE 157
QY 90 LKEAYLS-KTNRIKQKITNEMFIQMTQPIYDSLMMVDRGLGIYNPNNEVFALVRARG-- 146
DB 158 LLSLFLSEKVDRIELVTRFVSLVSSRFVITQLPLDTQG--LEAADDEIFRLTTRGGQF 215
QY 147 -----FDKALS--EGLHKSMDNQ-----AVSILVAKVEEIF 177
DB 216 QVERQTTSQARPLPRDMIFEQDPVQILDLSLLPLYSNQLLRALQESAELARVTAMS 275

QY 178 KDSVNYGDV 186

DB 276 NASENAGEL 284

RESULT 5

EX7L_FUSNN STANDARD; PRT; 404 AA.
ID Q8REN3;
AC Q8REN3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE (Exonuclease VII large subunit).
GN XSEA OR FN1066.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Fusch G., Haselkorn R.,
RA Fontein M., Kyrpides N., Overbeek R.,
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
CC -!- FUNCTION: Bidirectionally degrades single-stranded DNA into large
acid-insoluble oligonucleotides, which are then degraded further
into small acid-soluble oligonucleotides (by similarity).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5' to 3'-
or 3' to 5'-direction to yield nucleoside 5'-phosphates.
CC -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

-!- SIMILARITY: BELONGS TO THE XSEA FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE010612; AAL95262.1; -
HAMAP; MF_00378; 1.
InterPro; IPR003753; Exonuc_VII_L.
InterPro; IPR004365; tRNA_anti.
PIfam; PF02601; Exonuc_VII_L; 1.
PIfam; PF01336; tRNA_anti; 1.
TIGRFAMs; TIGR00237; xseA; 1.
Hydrolase; Nuclease; Exonuclease; Complete proteome.
SEQUENCE 404 AA; 46462 MW; 9541A9D8F0E1A38D CRC64;

Query Match 9.3%; Score 91; DB 1; Length 404;
Best Local Similarity 24.1%; Pred. No. 4;
Matches 39; Conservative 44; Mismatches 47; Indels 32; Gaps 10;

QY 32 QNRSLQDVPVYMLQNRSEYITQGVDSHIVDGKTEEIEKIATKRAIRVAQIVHVK 91
DB 264 EKESLEB-----LKAREIYITKLLKS--YVDSMKRELLLEI--ENYLYKNFPTVASLR 314
QY 92 EAYLSKTNRIKQKITNEMFIQMTQPIYDSLMMVDRGLGIYNPNNEVFALVRARGFD--- 148
DB 315 ESIVEKEIQLKEAM--ESFIRQKRNIFEN--KIDKISV-LNPIN-----TLKRGYTVSQ 363

QY 149 -KDALSEGLHKSMDNQAVSIL-----VAKVEEIPKDSVN 182

DB 364 VKNKRIDVLDIDIEINDEMTILDKGVISVVK-EKIYEKNIN 404

RESULT 6

NEBU_HUMAN STANDARD; PRT; 6669 AA.
ID NEBU_HUMAN STANDARD; PRT; 6669 AA.
AC P20929; Q15346;
DT 01-FEB-1991 (Rel. 17, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nebulin.
GN NEB.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95257391; PubMed=7739042;
RA Labelit S., Kolmerer B.;
RT "The complete primary structure of human nebulin and its correlation
to muscle structure.";
RL J. Mol. Biol. 248:308-315(1995).
RN [2]
RP PARTIAL PRELIMINARY SEQUENCE.
RX MEDLINE=88284704; PubMed=3397062;
RA Zeviani M., Darras B.T., Rizzuto R., Salvati G., Betto R.,
RA Bonilla E., Miranda A.F., Du J., Samitt C., Dickson G., Walsh F.S.,
RA Dimauro S., Francke U., Schon E.A.;
RT "Cloning and expression of human nebulin cDNAs and assignment of the
gene to chromosome 2q31-q32.";
RL Genomics 2:249-256(1988).
RN [3]
RP STRUCTURE BY NMR OF 6610-6669.
RX MEDLINE=98179559; PubMed=9514727;
RA Politou A.S., Millevol S., Gautel M., Kolmerer B., Pastore A.;
RT "SH3 in muscles: solution structure of the SH3 domain from nebulin.";
RL J. Mol. Biol. 276:189-202(1998).
CC -!- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MAINTAINING

CC THE STRUCTURAL INTEGRITY OF SARCOMERES AND THE MEMBRANE SYSTEM
CC ASSOCIATED WITH THE MYOFIBRILS. BIND AND STABILIZE F-ACTIN.
CC -!- TISSUE SPECIFICITY: MUSCLE-SPECIFIC. LOCATED IN THE THIN FILAMENT
CC OF STRIATED MUSCLE.
CC
CC -!- DISEASE: Defects in NEB are a cause of the autosomal recessive
CC form of nemaline myopathy (NEM2).
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -!- SIMILARITY: Contains 178 nebulin repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC ENBL; X83957; CAA58788.1; ..
CC ENBL; M19668; AAA59916.1; ALT_SEQ.
CC ENBL; M19669; AAA59917.1; ALT_SEQ.
CC PIR; S55024; S55024.
CC PDB; 1ARX; 28-JAN-98.
CC PDB; 1NEB; 24-DEC-97.
CC Genew; HGNC:7720; NEB.
CC MIM; 161650; ..
CC MIM; 256030; ..
CC GO; GO:0015629; C:actin cytoskeleton; TAS.
CC GO; GO:0030017; C:sarcomere; NAS.
CC GO; GO:0003792; F:regulation of actin thin filament length ac. . .; NAS.
CC GO; GO:0008307; F:structural constituent of muscle; TAS.
CC GO; GO:0007525; P:somatic muscle development; NAS.
CC InterPro; IPR000900; Nebulin.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00880; Nebulin; 146.
CC PRINTS; PF00018; SH3; 1.
CC PRINTS; PR00510; NEBULIN.
CC PRINTS; PR00452; SH3DOMAIN.
CC ProDom; PD000066; SH3; 1.
CC PROSITE; PS00002; SH3; 1.
KW Actin-binding; Muscle protein; Cytoskeleton; Repeat; 3D-structure.
FT REPEAT 76 107
FT REPEAT 113 143 NEBULIN 1.
FT REPEAT 148 178 NEBULIN 2.
FT REPEAT 183 213 NEBULIN 3.
FT REPEAT 218 248 NEBULIN 4.
FT REPEAT 253 283 NEBULIN 5.
FT REPEAT 289 318 NEBULIN 6.
FT REPEAT 324 354 NEBULIN 7.
FT REPEAT 363 393 NEBULIN 8.
FT REPEAT 398 428 NEBULIN 9.
FT REPEAT 434 464 NEBULIN 10.
FT REPEAT 502 532 NEBULIN 11.
FT REPEAT 537 567 NEBULIN 12.
FT REPEAT 573 603 NEBULIN 13.
FT REPEAT 611 641 NEBULIN 14.
FT REPEAT 681 711 NEBULIN 15.
FT REPEAT 749 779 NEBULIN 16.
FT REPEAT 784 814 NEBULIN 17.
FT REPEAT 820 850 NEBULIN 18.
FT REPEAT 858 888 NEBULIN 19.
FT REPEAT 893 923 NEBULIN 20.
FT REPEAT 924 954 NEBULIN 21.
FT REPEAT 959 990 NEBULIN 22.
FT REPEAT 993 1023 NEBULIN 23.
FT REPEAT 1028 1058 NEBULIN 24.
FT REPEAT 1064 1094 NEBULIN 25.
FT REPEAT 1102 1132 NEBULIN 26.
FT REPEAT 1137 1167 NEBULIN 27.
FT REPEAT 1168 1198 NEBULIN 28.
FT REPEAT 1204 1234 NEBULIN 29.
FT REPEAT 1237 1267 NEBULIN 30.
FT REPEAT 1272 1302 NEBULIN 31.
FT REPEAT 1308 1338 NEBULIN 32.
FT REPEAT 1338 1338 NEBULIN 33.
FT REPEAT 1346 1376 NEBULIN 34.
FT REPEAT 1381 1411 NEBULIN 35.
FT REPEAT 1412 1442 NEBULIN 36.
FT REPEAT 1448 1478 NEBULIN 37.
FT REPEAT 1481 1511 NEBULIN 38.
FT REPEAT 1516 1546 NEBULIN 39.
FT REPEAT 1552 1582 NEBULIN 40.
FT REPEAT 1580 1620 NEBULIN 41.
FT REPEAT 1625 1655 NEBULIN 42.
FT REPEAT 1656 1686 NEBULIN 43.
FT REPEAT 1692 1722 NEBULIN 44.
FT REPEAT 1725 1755 NEBULIN 45.
FT REPEAT 1760 1790 NEBULIN 46.
FT REPEAT 1796 1826 NEBULIN 47.
FT REPEAT 1834 1864 NEBULIN 48.
FT REPEAT 1869 1899 NEBULIN 49.
FT REPEAT 1900 1930 NEBULIN 50.
FT REPEAT 1936 1966 NEBULIN 51.
FT REPEAT 1969 1999 NEBULIN 52.
FT REPEAT 2004 2034 NEBULIN 53.
FT REPEAT 2040 2070 NEBULIN 54.
FT REPEAT 2078 2108 NEBULIN 55.
FT REPEAT 2113 2143 NEBULIN 56.
FT REPEAT 2144 2174 NEBULIN 57.
FT REPEAT 2180 2210 NEBULIN 58.
FT REPEAT 2213 2243 NEBULIN 59.
FT REPEAT 2248 2278 NEBULIN 60.
FT REPEAT 2284 2314 NEBULIN 61.
FT REPEAT 2322 2352 NEBULIN 62.
FT REPEAT 2357 2387 NEBULIN 63.
FT REPEAT 2388 2418 NEBULIN 64.
FT REPEAT 2423 2453 NEBULIN 65.
FT REPEAT 2456 2486 NEBULIN 66.
FT REPEAT 2491 2521 NEBULIN 67.
FT REPEAT 2527 2557 NEBULIN 68.
FT REPEAT 2565 2595 NEBULIN 69.
FT REPEAT 2600 2630 NEBULIN 70.
FT REPEAT 2631 2661 NEBULIN 71.
FT REPEAT 2666 2696 NEBULIN 72.
FT REPEAT 2699 2729 NEBULIN 73.
FT REPEAT 2734 2764 NEBULIN 74.
FT REPEAT 2800 2838 NEBULIN 75.
FT REPEAT 2843 2873 NEBULIN 76.
FT REPEAT 2874 2904 NEBULIN 77.
FT REPEAT 2909 2939 NEBULIN 78.
FT REPEAT 2942 2972 NEBULIN 79.
FT REPEAT 2977 3007 NEBULIN 80.
FT REPEAT 3013 3043 NEBULIN 81.
FT REPEAT 3051 3081 NEBULIN 82.
FT REPEAT 3086 3116 NEBULIN 83.
FT REPEAT 3117 3147 NEBULIN 84.
FT REPEAT 3152 3182 NEBULIN 85.
FT REPEAT 3200 3230 NEBULIN 86.
FT REPEAT 3220 3250 NEBULIN 87.
FT REPEAT 3256 3286 NEBULIN 88.
FT REPEAT 3294 3324 NEBULIN 89.
FT REPEAT 3329 3359 NEBULIN 90.
FT REPEAT 3360 3390 NEBULIN 91.
FT REPEAT 3395 3425 NEBULIN 92.
FT REPEAT 3428 3458 NEBULIN 93.
FT REPEAT 3463 3493 NEBULIN 94.
FT REPEAT 3499 3529 NEBULIN 95.
FT REPEAT 3537 3567 NEBULIN 96.
FT REPEAT 3572 3602 NEBULIN 97.
FT REPEAT 3603 3633 NEBULIN 98.
FT REPEAT 3638 3668 NEBULIN 99.
FT REPEAT 3671 3701 NEBULIN 100.
FT REPEAT 3706 3736 NEBULIN 101.
FT REPEAT 3742 3772 NEBULIN 102.
FT REPEAT 3780 3810 NEBULIN 103.
FT REPEAT 3815 3845 NEBULIN 104.
FT REPEAT 3846 3876 NEBULIN 105.
FT REPEAT 3846 3876 NEBULIN 106.

```

DR HSP: O63245; 2HPH.
DR WormPep; F26A1.2; CF02683.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
DR KW Hypothetical protein; DNA-binding; Nuclear protein;
KW Transcription regulation.
FT DNA_BIND 171 262
SQ SEQUENCE 461 AA; 54171 MW; 3FCFEDE123297370 CRC64;

Query Match          9.1%; Score 89; DB 1; Length 461;
Best Local Similarity 24.2%; Pred. No. 6.8;
Matches 30; Conservative 23; Mismatches 47; Indels 24; Gaps 4

QY      84 QNIVHKLKAYLSKTNRIKQIKNEMFIQTQTFYDSLM-----NVDRLGIYINPNNEVVF 139
Db       318 RKTIPKLETRVLEHWQMEHKATPEEFEEFVPTSYSLSCVENISYSGEGVGPNGNDELY 377

QY      140 AL-----VRARGFKDALSGLHWKSLDNQAVSLIVAKV-----RIFKDSVNY 183
Db       378 DLSDRQIICFHNIRA---YHSIQKCKHKNYSNWTPLSLIVQEAIVKDIVEDDIN 433

QY      184 GDVK 187
Db       434 GTIE 437

RESULT 8
SMC4_YEAST
ID _SMC4_YEAST STANDARD; PRT; 1418 AA.
AC Q12257;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Structural maintenance of chromosome 4.
GN SMC4 OR YLR086W OR IY449.5.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_taxid=4932;
RN [1]
RP SEQUENCE FROM N/A.
RC STRAIN=3288C / AB972;
RX MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansorge W.,
RA Baner V., Bruckner M., Delius H., Dubois E., Duesterhoeft A.,
RA Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA Louis E.-J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,
RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetelle D., Purnelle B., Reichmann S., Rieger M., Rinke M., Rose M.,
RA Scharte M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urestarazu L.A., Vandenbol M., Verhasselt P.,
RA Vierendeels F., Voet M., Volktaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hobeisel J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL Nature 387:87-90(1997).
RN [2]
RP IDENTIFICATION IN A CONDENSIN COMPLEX WITH SMC2; BRN1; YCS4 AND
RP YG1.
RX MEDLINE=20273907; PubMed=10811823;
RA Freeman L., Aragon-Alcaide L., Strunnikov A.;
RT "The condensin complex governs chromosome condensation and mitotic
RT transmission of rDNA.";
RL J. Cell Biol. 149:811-824(2000).
CC -I- FUNCTION: Central component of the condensin complex, a complex
CC required for conversion of interphase chromatin into mitotic-like
CC condense chromosomes. The condensin complex probably introduces
CC positive supercoils into relaxed DNA in the presence of type I

```

CC topoisomerases and converts nicked DNA into positive knotted forms
 CC in the presence of type II topoisomerases.
 CC -1- SUBUNIT: Forms an heterodimer with SMC2. Component of the
 CC condensin complex, which contains the SMC2 and SMC4 heterodimer,
 CC and three non-SMC subunits that probably regulate the complex.
 CC BRN1, YCS4 and YCG1/YCS5.
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic. In interphase
 CC cells, the majority of the condensin complex is found in the
 CC cytoplasm, while a minority of the complex is associated with
 CC chromatin. A subpopulation of the complex however remains
 CC associated with chromosome foci in interphase cells. During
 CC mitosis, most of the condensin complex is associated with the
 CC chromatin. At the onset of prophase, condensin associates with
 CC chromosome arms and to chromosome condensation. Dissociation from
 CC chromosomes is observed in late telophase.
 CC -1- DOMAIN: The flexible hinge domain, which separates the large
 CC intramolecular coiled coil regions, allows the heterodimerization
 CC with SMC2, forming a V-shaped heterodimer (By similarity).
 CC -1- SIMILARITY: Belongs to the SMC family. SMC4 subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z73258; CAA97646.1; -
 CC EMBL; Z73259; CAA97648.1; -
 CC EMBL; U53880; AAB67590.1; -
 CC PIR; S64918; S64918.
 CC SGD; S0004076; SMC4.
 CC GO; GO:000576; Chromatin complex; IPI.
 CC GO; GO:0007076; Pmitotic chromosome condensation; IMP.
 CC InterPro; IPR003405; SMC_C.
 CC InterPro; IPR003395; SMC_N.
 CC Pfam; PF02483; SMC_C; 1.
 CC Pfam; PF02463; SMC_N; 1.
 CC DNA condensation; Mitosis; Cell cycle; ATP-binding; Coiled coil;
 CC Nuclear protein.
 CC NP-BIND 185 192 ATP (POTENTIAL).
 CC FT DOMAIN 345 673 COILED COIL (POTENTIAL).
 CC FT DOMAIN 674 848 FLEXIBLE HINGE.
 CC FT DOMAIN 849 1172 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1224 1263 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1323 1358 ALA/ASP-RICH (DA-BOX).
 CC SEQUENCE 1418 AA; 162188 MW; F0E6B72F8BFD374 CRC64;
 CC -----
 CC Query Match 9.0%; Score 87.5; DB 1; Length 1418;
 CC Best Local Similarity 21.6%; Pred. No. 34;
 CC Matches 37; Conservative 32; Mismatches 61; Indels 41; Gaps 6;
 CC -----
 CC QY 55 GVDSSHIVDGKTEETETK-----ATKRAIVRAQNVHKL-----FAYLSK 97
 CC Db 836 GTNQSQKVDYTPPEVDKTELSELRNFRVSDTVEHEBELKKLRDHPDLESQISK 895
 CC QY 98 TNRKIKQTNEMFIQNTQ-----PIYDS-----LMNVDRLL-GYIYNPNNEEYFA 140
 CC Db 896 AENEADSLASLTLAQVKEAEYAVKAVSKAQLNVVMKLERLGRYNDLQSETTKT 955
 CC QY 141 LVPARGFDKDALSEGHLKMSLDN-----QAVSLVAKVEIFKDSVNYG 184
 CC Db 956 KEIKGLQDEIMKIGGKIQMQNSKVESVCQKLDILVAKLKKVKSASKSG 1006
 CC -----
 CC RESULT 9
 CC MYH8_HUMAN
 CC ID MYH8_HUMAN STANDARD; PRT; 1937 AA.
 CC AC P13535; Q14910;
 CC DT 01-JAN-1990 (Rel. 13, Created)
 CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Myosin heavy chain, skeletal muscle, perinatal (MYHC-perinatal).
 GN MYH8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=90123631; PubMed=2373371;
 RA Karsch-Mirachi I., Peghali R., Shows T.B. Jr., Leinwand L.A.;
 RT "Generation of a full-length human perinatal myosin heavy-chain-
 RT encoding cDNA.";
 RL Gene 89:289-294(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=95324556; PubMed=7601129;
 RA Jullian E.H., Kelly A.M., Pompidou A.J., Hoffman R., Schiaffino S.,
 RA Stedman H.H., Rubinstein N.A.;
 RT "Characterization of a human perinatal myosin heavy-chain
 RT transcript.";
 RL Eur. J. Biochem. 230:1001-1006(1995).
 RN [3]
 RP SEQUENCE OF 502-1937 FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=90235862; PubMed=1691980;
 RA Bober B., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
 RA Arnold H.H.;
 RT "Identification of three developmentally controlled isoforms of human
 RT myosin heavy chains.";
 RL Eur. J. Biochem. 189:55-65(1990).
 RN [4]
 RP SEQUENCE OF 860-1937 FROM N.A.
 RX MEDLINE=89234168; PubMed=2715179;
 RA Peghali R., Leinwand L.A.;
 RT "Molecular genetic characterization of a developmentally regulated
 RT human perinatal myosin heavy chain.";
 RL J. Cell Biol. 108:1791-1797(1989).
 RN [5]
 RP SEQUENCE OF 1-46 FROM N.A.
 RA Esser K., Tidhar A., Myszkowski M.;
 RT "Isolation and characterization of the human perinatal MHC promoter.";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HWM). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -1- SIMILARITY: Contains 1 IQ domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M36769; AAC17185.1; -
 CC EMBL; Z38133; CAA86293.1; -
 CC EMBL; X51592; CAA35941.1; -
 CC EMBL; AF067143; AAC21557.1; -

```

OX NCBI_TaxID=55529;
RN [1]
RP SEQUENCE FROM N.A.
RR Leitsch C.E.W., Kowalik K.V., Douglas S.E.;
RT "The atpA gene cluster of a cryptomonad, Guillardia theta: a piece in
RA the puzzle of chloroplast genome development.";
RNL J. Phycol. 35:128-135(1999).
RN [2]
RP SEQUENCE FROM N.A.
RR MEDLINE=99128221; PubMed=9929392;
RX Douglas S.E., Penny S.L.;
RA "The plastid genome of the cryptophyte alga, Guillardia theta:
RT complete sequence and conserved syntenic groups confirm its common
RT ancestry with red algae.";
RNL J. Mol. Evol. 48:236-244(1999).
RC -!- FUNCTION: Associates with the EF-Tu.GDP complex and induces the
CC exchange of GDP to GTP. It remains bound to the aminoacyl-tRNA.EF-
CC Tu.GTP complex up to the GRP hydrolysis stage on the ribosome
CC (by similarity).
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: BELONGS TO THE EF-TS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CR EMBL; AF041468; AAC35672.1; -.
DR HSSP; P43895; 1AIP.
DR HAMAP; MF_00050; -.
DR InterPro; IPR001816; EF_TS.
DR InterPro; IPR000449; UBA_domain.
DR Pfam; PF00889; EF_TS; 1.
DR Pfam; PF00627; UBA; 1.
DR TIGRFAMs; TIGR00116; tsf; 1.
DR PROSITE; PS01126; EF_TS_1; 1.
DR PROSITE; PS01127; EF_TS_2; 1.
DR KW Elongation factor; Protein biosynthesis; Chloroplast.
DR QY SEQUENCE 219 AA; 24998 MW; 36E662FC83130049 CRC64;
SQ
Query Match 8.8%; Score 85.5; DB 1; Length 219;
Best Local Similarity 23.9%; Pred. No. 5.3;
Matches 39; Conservative 21; Mismatches 56; Indels 47; Gaps 6;
QY 20 LMLAQCTCPNTSQNSFLQDVPVWVLQRSEVITQGVDSHIVDGKTEIEKIATKEAT 79
DB 99 IAMQIACPNDVIKT--SDIPNEIIQKEKEMNKND---LDNKPTEIKEIVEGR-- 150
QY 80 IRVAQNIVHKLEAYSKTRIRIKOKITNMFTQMTOPIYDSLNVDRGLGINPNNEVF 139
DB 151 -----IQKLKSLMDQSIVR-----DSSLIEEL---IKENIAKLG 185
QY 140 ALVRARGFDKDALSEGHLHKMSLDAQSVLVAKVEIFKDSVN 182
DB 186 ENIQRRFRFTLGEGLEKR-----EDNFNEEVN 214
RESULT 11
RA25 YEAST
ID RA25 YEAST STANDARD; PRT; 843 AA.
AC Q00578;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE DNA repair helicase RAD25.
GN RAD25 OR SSL2 OR UVS112 OR YIL143C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI TaxID=4932; OX
```

RA25 YEAST	RA25 YEAST	STANDARD;	PRT;	843 AA.
ID	RA25_YEAST			
DT	01-APR-1993	(Rel. 25, Created)		
DT	01-APR-1993	(Rel. 25, Last sequence update)		
DT	15-SEP-2003	(Rel. 42, Last annotation update)		
DE	DNA repair helicase RAD25.			
GN	RAD25 OR SSL2 OR UVS112 OR YIL143C.			
OC	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
NCBI	TaxID=4932;			
OX	NCBI TaxID=4932;			

RN RP SEQUENCE FROM N.A.
RC STRAIN=9288C;
RX MEDLINE=92298384; PubMed=1318786;
RA Gulyas K.D., Donahue T.F.;
RT "SSL2, a suppressor of a stem-loop mutation in the HIS4 leader
RL encodes the yeast homolog of human ERCC-3";
RN Cell 69:1031-1042(1992).
[2]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=93087538; PubMed=1333609;
RA Park E.K., Gunder S.N., Weeda G., Hoeijmakers J.H., Prakash S.,
RA Prakash L.;
RT "RAD25 (SSL2), the yeast homolog of the human Xeroderma pigmentosum
RL group B DNA repair gene, is essential for viability";
RN Proc. Natl. Acad. Sci. U.S.A. 89:11416-11420(1992).
[3]
RN RP SEQUENCE FROM N.A.
RC STRAIN=9288C / AB972;
RX PubMed=9169870;
RA Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,
RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
RA Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,
RA Moulé S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
RA Skelton J., Smith V., Walsh S., Whitehead S., Barrett B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX";
RN Nature 387:84-87(1997).
CC -1- FUNCTION: PROBABLY AN ATP-DEPENDENT DNA HELICASE INVOLVED IN
CC EXCISION REPAIR OF DNA. MAY HAVE A DNA UNWINDING FUNCTION. MAY
CC HAVE A NONESSENTIAL FUNCTION IN NUCLEOTIDE EXCISION REPAIR AND
CC AN ESSENTIAL FUNCTION IN TRANSLATION INITIATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- MISCELLANEOUS: A C-TERMINAL DELETION RENDERS YEAST HYPERSENSITIVE
CC TO UV LIGHT.
CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. RAD25/XPB SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z38059; CAA86135.1; -
CC DR EMBL; M94176; AAA35102.1; -
CC DR EMBL; L01414; AAA34942.1; -
CC DR PIR; S31272; S31272.
CC TRANSFAC; T02192; -
CC SGD; S0001405; SSJ2.
CC DR InterPro; IPR001410; DEAD.
CC DR InterPro; IPR001650; Helicase C.
CC DR InterPro; IPR001161; XPB DNA repair.
CC DR Pfam; PF00271; helicase_C_1
CC DR PRINTS; PR00851; XRODRMGWNTB.
CC DR SMART; SM00487; DEXDC; 1.
CC DR SMART; SM00490; HELIC; 1.
CC DR TIGRFAMS; TIGR00603; rad25; 1.
CC DR Helicase; DNA repair; ATP-binding; Nuclear protein.
CC KW DOMAIN 64 75 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 302 309 ASP/GLU-RICH (ACIDIC).
FT NP_BIND 386 393 ATP (BY SIMILARITY).
FT SITE 487 491 DEVH BOX.
FT MUTAGEN 427 427 W->L: SUPPRESSOR MUTANT.
FT CONFLICT 9 9 P -> S (IN REF. 2).
FT CONFLICT 48 48 S -> L (IN REF. 2).
SQ SEQUENCE 843 AA; 95340 MW; FA4013B8156FE1C5 CRC64;

Query Match 8.7%; Score 85; DB 1; Length 843;
Best Local Similarity 22.1%; Pred. No. 29;
Matches 33; Conservative 33; Mismatches 45; Indels 36; Gaps 7;
QY 64 GKTEIEIKATKTRATIRVAQNIHVHKLKEAYLSKTNRIKQKIT-NEMFIQMTQ-PIYDSL 121

Db 187 GLETTDDIISVLDRLSKVPFAESIINFIKGATIS-YGKVKLVIKNRYFVEVTDQADILQML 245
QY 122 MNDRLG-----IVINPNN-EYVFALV-----RARGF 147
Db 246 LNDVIGPLRIDSQHQVQPPEDVLQQLQQTAGKPAITNVNPDVEAVFSVAVIGDNEREE 305
QY 148 DKDALSEGLHKMSLDNQAVSILVAKVBEI 176
Db 306 EDDDI-DAVHSFEIANESVEVVKRCQEI 333

RESULT 12
Y307_MYCPN
ID Y307_MYCPN STANDARD; PRT; 1244 AA.
AC P75342;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein MG307 precursor (A05_orf1244).
GN MPN436 OR MP405.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Hermann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
CC pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE MG307 / MG338 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000039; AAB96053.1; -
CC DR PIR; S73731; S73731.
CC DR PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
CC KW Hypothetical protein; Lipoprotein; Membrane; Signal;
CC Complete proteome.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 1244 HYPOTHETICAL LIPOPROTEIN MG307 HOMOLOG.
FT LIPID 28 28 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 1244 AA; 139117 MW; BDAFIAED587882A CRC64;

Query Match 8.7%; Score 85; DB 1; Length 1244;
Best Local Similarity 21.1%; Pred. No. 46;
Matches 54; Conservative 44; Mismatches 76; Indels 82; Gaps 13;
QY 4 KSFLKSKOLFCLGLG-----VLMLOACTCPNTS-----QRNS----- 35
Db 2 KKFLKPKQFWLLTGLGGFLSTSVILAAACATPSNSALQTVFKARSSQFFNGBEGGSQSALT 61
QY 36 -----FLQDVPYWMQNRSEYITQGV-DSSHIVDGKTEIEKI--ATKR 77
Db 62 ALKNPVANKQFIAAPLLKALEAWYNNEDKKITQFLDKTSNVDSQVTTAVDKVVSARN 121
QY 78 ATIRVAQNI-----HKLKEAYLSK-TNRIKQKITNEMFIQ-----MTQPI-Y 118
Db 122 KSLFVQDQLDNDAGGSEATWKAQKLEQLISDFASRVFQK--NYLNTKKQGVSTGFTY 179
QY 119 DSL-----MNVDRLGIVINPNNNEVEFALVRARFDK-----DALSEGLHKMSLDNQ 165

Db 180 DELHESKWNKFSAPRSETNDFFAKIQSVFQDQVETDPTLISQVNYKYSAPSQ 239

QY 166 VSLIVKVEEIFKDSV 181

Db 240 LGQIVNR--EKLKDKL 253

RESULT 13

MYH4_RABIT STANDARD; PRT; 1938 AA.

AC Q28641;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Myosin heavy chain, skeletal muscle, juvenile.

OS Oryctolagus cuniculus (Rabbit)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI_TaxID=9986;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=New Zealand white; TISSUE=Skeletal muscle;

RA Maeda K., Hostinova E., Roesch-Kleinlauf A., Schuster H., Gasperik J.,

RT Wittinghofer A.;

RT Isolation, sequencing of myosin heavy chain cDNA from rabbit

RT skeletal muscle and a novel cosynthesis of S-1 fragment with the

RT essential and regulatory light chains.";

RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: MUSCLE CONTRACTION.

CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2

CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)

CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).

CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.

CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING

CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,

CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.

CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT

CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE

CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED

CC SUBFRAGMENT (S2).

CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.

CC -!- SIMILARITY: Contains 1 IQ domain.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; U32574; AAA74199.1; .

DR PIR; A59293; A59293.

DR HSSP; P13338; 2MYS.

DR InterPro; IPR000048; IQ region.

DR InterPro; IPR001609; myosin head.

DR InterPro; IPR004009; Myosin N.

DR InterPro; IPR002928; Myosin tail.

DR Pfam; PF00612; IQ; 2.

DR Pfam; PF00063; myosin head; 1.

DR Pfam; PF02736; Myosin_N; 1.

DR Pfam; PF01576; Myosin_tail; 1.

DR PRINTS; PR00193; MYOSINHEAVY.

DR ProDom; PD000355; myosin_head; 1.

DR SMART; SM00015; IQ; 1.

DR SMART; SM00242; MYSC; 1.

DR PROSITE; PS50096; IQ; 1.

DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;

KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;

KW Multigene family.

FT DOMAIN 1 783 MYOSIN HEAD-LIKE.

FT DOMAIN 784 813 IQ.

FT DOMAIN 842 1938 COILED COIL (POTENTIAL).

FT NP_BIND 179 186 ATP (POTENTIAL).

FT DOMAIN 658 680 ACTIN-BINDING (BY SIMILARITY).

FT DOMAIN 760 774 ACTIN-BINDING (BY SIMILARITY).

FT MOD_RES 35 35 METHYLATION (MONO-) (BY SIMILARITY).

FT MOD_RES 130 130 METHYLATION (TRI-) (BY SIMILARITY).

FT MOD_RES 552 552 METHYLATION (TRI-) (BY SIMILARITY).

FT MOD_RES 756 756 METHYLATION (TRI-) (BY SIMILARITY).

FT MOD_RES 698 698 ALKYLATION (SH-1) (BY SIMILARITY).

FT MOD_RES 708 708 ALKYLATION (SH-2) (BY SIMILARITY).

SQ SEQUENCE 1938 AA; 223064 MW; D8A9A2EC5B182626 CRC64;

Query Match 8.7%; Score 85; DB 1; Length 1938;

Best Local Similarity 26.3%; Pred. No. 78;

Matches 44; Conservative 31; Mismatches 42; Indels 50; Gaps 11;

QY 31 SORNFLODPVYMWLQNRSEYITQGVDSHIVDGKKTTEEIKIATKATIRVAQNIYH-- 88

Db 1365 SKANS---EVAQW---RTKYETDAI-----QRTTELEE-AKKLAQLQDAEEHVE 1408

QY 89 --KLKEAVLSKINRIKQKITNEMFTQTPYVDSIMNVDRIGYINPNNEVPAL-VRAR 145

Db 1409 AVNAKASLEKT---KQRLQNE-----VEDLMIDVER-----TNACALDKKQR 1450

QY 146 GFQDKALSEGLHK-----MSLDNQAVSILVAKVEEIFKDSVN 182

Db 1451 NFDK-ILAEWKHKYETHAELEASQKESRSLSTEVFKVKNAYEESLD 1496

RESULT 14

Y114_METJA STANDARD; PRT; 303 AA.

AC Q57578;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein M00114.

GN M00114.

OS Methanococcus jannaschii.

OC Archaea; Euryarchaeota; Methanococci; Methanococcales;

OC Methanocaldococcaceae; Methanocaldococcus.

OX NCBI_TaxID=2190;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;

RX MEDLINE=96337999; PubMed=8688087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,

RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,

RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,

RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,

RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,

RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,

RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,

RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;

RT "Complete genome sequence of the methanogenic archaeon, Methanococcus

RT jannaschii.";

RL Science 273:1058-1073(1996).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; U67469; AAB98110.1; .

DR PIR; B64314; B64314.

DR TIGR; M0114; .

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 303 AA; 34921 MW; D2B37C090A5837DA CRC64;

Query Match 8.6%; Score 84; DB 1; Length 303;

Best Local Similarity 23.0%; Pred. No. 10;

Matches 32; Conservative 26; Mismatches 55; Indels 26; Gaps 4;
QY 65 KTEIEIKIATKRAIRVAQNIIVHKLKEAYLS-----KTNEIKQKITNEMFIQWTP 116
DB 52 KSLEIFSQIDPEFLIYQNGEKEVKSALNFKPESIVEISKLSLVFENEVLANSYND 111
QY 117 IYDSLMNVDR-----GIYNPNNEEVALVRAR--FDKALSEGHLKMSLDN 163
DB 112 VYKLTITNKVIFMKPFKEKNGVIVYKNGKEVFAVYFGRKTLFGKAISK-----LKT 166
QY 164 QAVSILVAKVEIEFKDSVN 182
DB 167 FAVSEIIAKIEKISNEELN 185

RESULT 15
ID GREATER CHLTR STANDARD; PRT; 715 AA.
AC O84641;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transcription elongation factor greA (Transcript cleavage factor greA).
GN GRE A OR CT636.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/Cx;
RX MEDLINE=9900809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.;
RA "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis";
RT Science 282:754-759(1998).

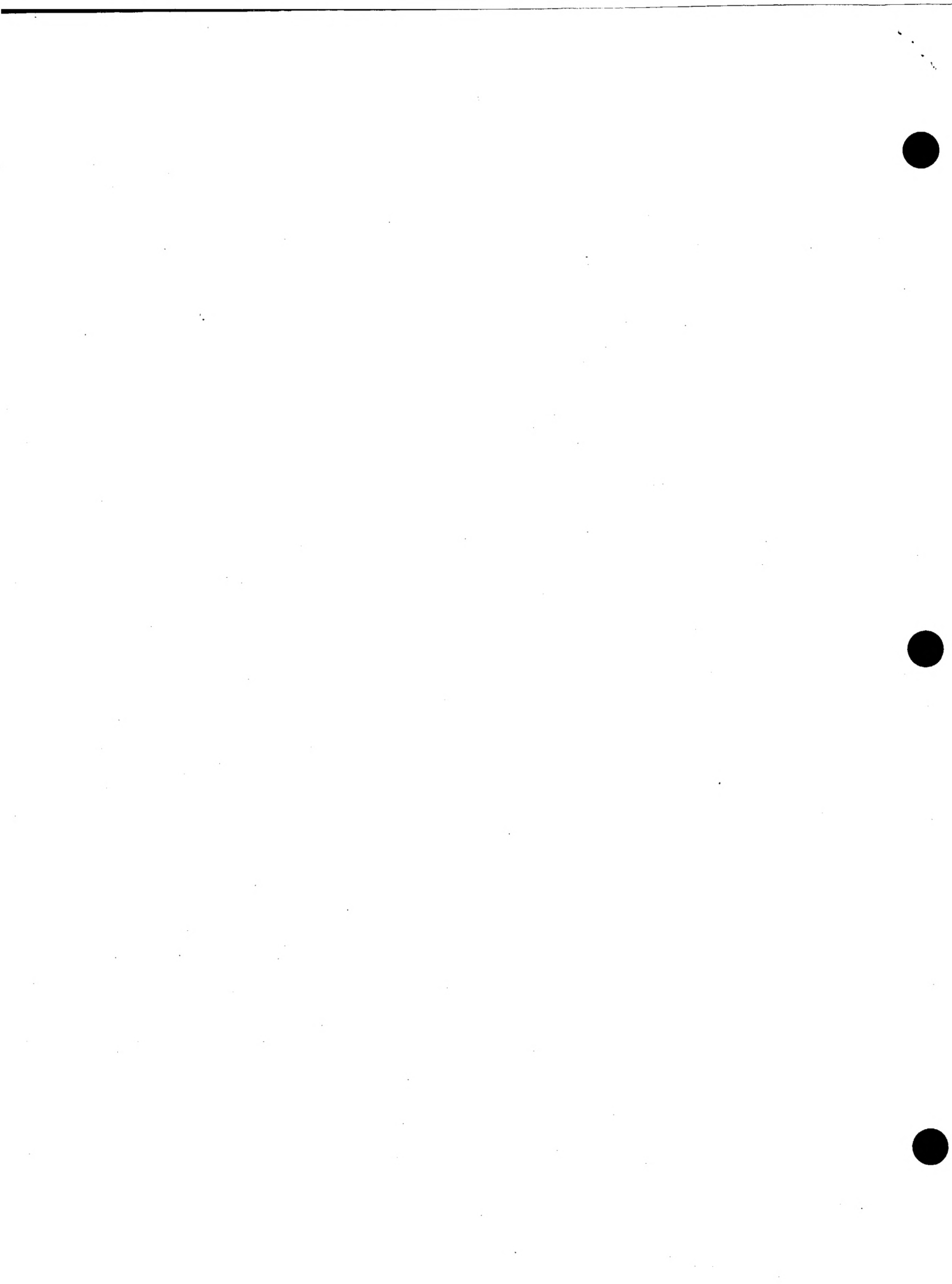
CC -!- FUNCTION: NECESSARY FOR EFFICIENT RNA POLYMERASE TRANSCRIPTION
CC ELONGATION PAST TEMPLATE-ENCODED ARRESTING SITES. THE ARRESTING
CC SITES IN DNA HAVE THE PROPERTY OF TRAPPING A CERTAIN FRACTION
CC OF ELONGATING RNA POLYMERASES THAT PASS THROUGH, RESULTING IN
CC LOCKED TERNARY COMPLEXES. CLEAVAGE OF THE NASCENT TRANSCRIPT BY
CC CLEAVAGE FACTORS SUCH AS GRE A OR GREB ALLOWS THE RESUMPTION OF
CC ELONGATION FROM THE NEW 3'-TERMINUS. GRE A RELEASES SEQUENCES OF
CC 2 TO 3 NUCLEOTIDES (BY SIMILARITY).
CC -!- SIMILARITY: Contains 1 GRAD1 (greA associated) domain.
CC -!- SIMILARITY: BELONGS TO THE GRE A/GREB FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AE001334; AAC68240.1; --
DR FIR; H71489; H71489.
DR HAMAP; MD_00001; fused; 1.
DR HAMAP; MF_00105; fused; 1.
DR InterPro; IPR006359; GreA.
DR InterPro; IPR001437; GreA_GreB.
DR Pfam; PF01272; GreA_GreB; 1.
DR Pfam; PF03449; GreA_GreB_N; 1.
DR TIGRfam; TIGR01462; greA; 1.
DR PROSITE; PS00829; GREAB_1; 1.
DR PROSITE; PS00830; GREAB_2; 1.
KW Transcription regulation; DNA-binding; Coiled coil; Complete proteome.
FT DOMAIN 1 502
FT DOMAIN 562 715
FT DOMAIN 603 641
FT SEQUENCE 715 AA; 80853 MW; 51F31448CBAF970 CRC64;
SQ

Query Match 8.5%; Score 82.5; DB 1; Length 715;
Best Local Similarity 24.1%; Pred. No. 38;
Matches 34; Conservative 28; Mismatches 62; Indels 17; Gaps 6;
QY 44 MLQNSEVITQGVDSHIVDGKTEIEIKIATKRAIRVAQNIIVHKLKEAYLSKTNRIKQ 103
DB 526 LLSSKCPQFTQG-DLGVL---RSLAEVVQPAKRGITPEEENILWTTSDSFTRNKKLQS 581
QY 104 KITNEMFIQWTPQPIYDLSLMNVDRIGIYNPNNEEVALVRARFPDKALSEGHLKMSLDN 163
DB 582 LVGKEM-VENAKEIEDARALGD-----LRENSVKIALER-----RARLQEEIHVLSEEI 630
QY 164 QAVSILVAKVEIEFKDSVNYG 184
DB 631 NRAKILTK-DAVFTDSVGVG 649

Search completed: December 10, 2003, 18:34:37
Job time : 12.5393 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:29:51 ; Search time 27.8953 Seconds
(without alignments)
1776.146 Million cell updates/sec

Title: US-10-080-113-3

Perfect score: 974

Sequence: 1 MLEKSFLLKSKQLFLCGLGLVL.....VEEIFKDSVNYGDKVKPIAM 192

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258032604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL 23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mnc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	974	100.0	192	16	O25318
2	932	95.7	192	2	Q9ZBD3
3	929	95.4	192	16	Q9ZLN5
4	99	10.2	3692	16	Q8CNU9
5	94	9.7	790	3	Q07732
6	93	9.5	682	5	Q9UAQ7
7	92.5	9.5	355	2	Q87224
8	92.5	9.5	2363	4	Q99968
9	91.5	9.4	268	5	Q8SVU2
10	91.5	9.4	404	2	Q9RNA5
11	91.5	9.4	450	4	Q9NW91
12	91.5	9.4	8094	5	Q8ILB9
13	91	9.3	5779	5	Q8IS80
14	90.5	9.3	1000	5	Q8IBX8
15	89.5	9.2	586	16	Q98RB1
16	89.5	9.2	993	5	Q8IJ59

17	89	9.1	791	16	Q9PQB3
18	88.5	9.1	672	16	Q9KH72
19	88.5	9.1	1071	10	Q8W0X3
20	88.5	9.1	1156	16	Q66878
21	88	9.0	1447	16	Q9PQB3
22	88	9.0	2129	5	Q8IAJ8
23	87.5	9.0	394	9	Q8AZK8
24	87.5	9.0	394	16	Q9CJB8
25	87.5	9.0	464	6	Q8WNT9
26	87.5	9.0	474	2	Q93CD6
27	87.5	9.0	477	2	Q8VLX1
28	87.5	9.0	489	16	Q932D7
29	87.5	9.0	699	4	Q9BWX0
30	87.5	9.0	710	12	Q91FQ5
31	87.5	9.0	1418	3	Q12267
32	87	8.9	287	16	Q8DIU9
33	87	8.9	1188	5	Q8I544
34	87	8.9	2391	5	Q8IKR4
35	86.5	8.9	434	16	Q8EDC0
36	86.5	8.9	669	10	Q9ZP55
37	86.5	8.9	698	4	Q8NEP0
38	86.5	8.9	698	4	Q9BZW7
39	86.5	8.9	1358	11	Q62411
40	86	8.8	397	2	Q8KX82
41	85.5	8.8	377	17	Q8TW22
42	85.5	8.8	649	4	Q96Q37
43	85.5	8.8	1686	5	Q8IFP9
44	85	8.7	218	2	Q8GMN6
45	85	8.7	266	16	Q8XJV4

ALIGNMENTS

RESULT 1

O25318 PRELIMINARY; PRT; 192 AA.
ID O25318
AC O25318;
DT 01-JAN-1998 (TRENDELrel. 05, Created)
DT 01-JAN-1998 (TRENDELrel. 05, Last sequence update)
DT 01-MAR-2002 (TRENDELrel. 20, Last annotation update)
DE Hypothetical protein HP0596.
GN HP0596.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]_TaxID=210;
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervatage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RT "The complete genome sequence of the gastric pathogen Helicobacter
pylori.";
RL Nature 388:539-547(1997).
DR EMBL; AB000573; AAD07665.1; -.
DR TIGR; HP0596; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 192 AA; 21878 MW; F643AFD91293D9CF CRC64;

Query Match 100.0%; Score 974; DB 16; Length 192;

Best Local Similarity 100.0%; Pred. No. 6.3e-70;

Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLEKSFLLKSKQLFLCGLGLVLQACTPNTSQNSFLQDVFYWLQNRSEYITQGVDSH 60

```

Db 1 MLEKSFLLSKQLFLCGLGVLMLQACTCPNTSQRSFLQDVPYMWLQNRSEYITQGVDSH 60
QY 61 IVDGKTEIEIKIATKRATIRVAQNIHVHKLKEAYLSKTNRIKQKITNEMFIQTQPIYDS 120
Db 61 IVDGKTEIEIKIATKRATIRVAQNIHVHKLKEAYLSKTNRIKQKITNEMFIQTQPIYDS 120
QY 121 LMNVDRLGIIYNPNNEEVFALVRARGFDKDALSEGHLKMSLDNQAVSILVAKVEEIPKDS 180
Db 121 LMNVDRLGIIYNPNNEEVFALVRARGFDKDALSEGHLKMSLDNQAVSILVAKVEEIPKDS 180
QY 181 VNYGDKVKPIAM 192
Db 181 VNYGDKVKPIAM 192

RESULT 2
Q9ZBD3 ID Q9ZBD3 PRELIMINARY; PRT; 192 AA.
AC Q9ZBD3;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE Superficial protein.
GN HPS.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=99081755; PubMed=9864228;
RA Yoshida M., Wakatsuki Y., Kobayashi Y., Itoh T., Murakami K.,
RA Mizoguchi A., Usui T., Chiba T., Kita T.;
RT "Cloning and characterization of a novel membrane-associated antigenic
RT protein of Helicobacter pylori.";
RL Infect. Immun. 67:286-293(1999).
DR EMBL; D30661; BAA28172.1; -
SQ SEQUENCE 192 AA; 21860 MW; 244DE262D2D535F19 CRC64;

Query Match 95.7%; Score 932; DB 2; Length 192;
Best Local Similarity 94.8%; Pred. No. 1.4e-66;
Matches 182; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 MLEKSFLLSKQLFLCGLGVLMLQACTCPNTSQRSFLQDVPYMWLQNRSEYITQGVDSH 60
Db 1 MLEKSFLLSKQLFLCGLGVLMLQACTCPNTSQRSFLQDVPYMWLQNRSAVITQGVDSH 60
QY 61 IVDGKTEIEIKIATKRATIRVAQNIHVHKLKEAYLSKTNRIKQKITNEMFIQTQPIYDS 120
Db 61 IVDGKTEIEIKIATKRATIRVAQNIHVHKLKEAYLSKSNRIKQKITNEMFIQTQPIFDS 120
QY 121 LMNVDRLGIIYNPNNEEVFALVRARGFDKDALSEGHLKMSLDNQAVSILVAKVEEIPKDS 180
Db 121 LMNVDRLGIIYNPNNEEVFALVRARGFDKDALSEGHLKMSLDNQAVSILVAKVEEIPKDS 180
QY 181 VNYGDKVKPIAM 192
Db 181 VNYGDKVKPIAM 192

RESULT 3
Q9ZLN5 ID Q9ZLN5 PRELIMINARY; PRT; 192 AA.
AC Q9ZLN5;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE Putative.
GN JHP0543.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.

```

```

OX NCBI_TaxID=85963;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merbuso D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
DR EMBL; AE001487; AAD06128.1; -
KW Complete proteome.
SQ SEQUENCE 192 AA; 21803 MW; 99945EB48CF4FB7B CRC64;

Query Match 95.4%; Score 929; DB 16; Length 192;
Best Local Similarity 94.3%; Pred. No. 2.4e-66;
Matches 181; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 1 MLEKSFLLSKQLFLCGLGVLMLQACTCPNTSQRSFLQDVPYMWLQNRSEYITQGVDSH 60
Db 1 MLEKSFLLSKQLFLCGLGVLMLQACTCPNTSQRSFLQDVPYMWLQNRSAVITQGVDSH 60
QY 61 IVDGKTEIEIKIATKRATIRVAQNIHVHKLKEAYLSKTNRIKQKITNEMFIQTQPIYDS 120
Db 61 IVDGKTEIEIKIATKRATIRVAQNIHVHKLKEAYLSKSNRIKQKITNEMFIQTQPIFDS 120
QY 121 LMNVDRLGIIYNPNNEEVFALVRARGFDKDALSEGHLKMSLDNQAVSILVAKVEEIPKDS 180
Db 121 LMNVDRLGIIYNPNNEEVFALVRARGFDKDALSEGHLKMSLDNQAVSILVAKVEEIPKDS 180
QY 181 VNYGDKVKPIAM 192
Db 181 VNYGDKVKPIAM 192

RESULT 4
Q8CNU9 ID Q8CNU9 PRELIMINARY; PRT; 3692 AA.
AC Q8CNU9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE FmtB protein.
GN SE1429.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AS016748; AAO05028.1; -
KW Complete proteome.
SQ SEQUENCE 3692 AA; 408837 MW; 7B5FF5B21DE246F9 CRC64;

Query Match 10.2%; Score 99; DB 16; Length 3692;
Best Local Similarity 19.2%; Pred. No. 91;
Matches 37; Conservative 44; Mismatches 72; Indels 40; Gaps 7;

QY 30 TSQRNSFLQDVPYMWLQNRSEYITQGVDSHIVDGK-----KTEEI 70
Db 2115 TEKNTAIQSIDDTTQAQRNNINGANTNALVDENLEDGKQLQRIVLSTQTKTQAKADIA 2174
QY 71 EKIAIKRATIRVAQNIHVHKLKEAYLSKTN-----RIKQKITNEMFIQTQPIYDSLM 122
Db 2175 QAIGQQRSTIDQNAATTEERKEALRLNQETNGVNDRIQAALANQVTDKNNILETIR 2234
QY 123 NVDRIGIYNPNNEEVFALVRARGFDKDALSEGHLKMSLDNQAVSILVAKVEEIPKDSVN 182
Db 123 NVDRIGIYNPNNEEVFALVRARGFDKDALSEGHLKMSLDNQAVSILVAKVEEIPKDSVN 182

```

```
Db 2235 NVEPI-VIVKPAKEI---IRKAAEQTLINQNDATLEEKQIAL--GKLEVKNEALN 2288
QY 183 Y-----GDKV 188
Db 2289 QVSOAHNNNDVKI 2301

RESULT 5
Q07732 PRELIMINARY; PRT; 790 AA.
ID Q07732
AC Q07732
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Chromosome IV reading frame ORF YDL239C.
GN ADY3 OR YDL239C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Alt-Noerbe J., Schneider C., Moro M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z74287; CA98849.1; -.
DR SGD; S0002398; ADY3.
SQ SEQUENCE 790 AA; 91739 MW; 6F8D43B78F38928F CRC64;

Query Match 9.7%; Score 94; DB 3; Length 790;
Best Local Similarity 21.8%; Pred. No. 35;
Matches 52; Conservative 41; Mismatches 79; Indels 66; Gaps 11;

QY 5 SFKSKQLFGLGLVLMQACTPNTSQ--RNSFLQDVPYWM---LQNRSEYITQGVDS 58
Db 433 SYQNLSLFEXDLGKFIEMKCGHSPMFQNGFAKLYPDFQDIKLENMEQY-KQLKKG 491
QY 59 SHIVDGKTEETEKIATKATIR-----VAQNVHKLKEAYLSKTRIKQKTNEMFTOM 113
Db 492 IELLEKNDRIREKIIISVFKLINERLHFVQOQSHKIK--YIQ-----KEALTKEQQFRL 544
QY 114 TOPIYDSLNV-----DRLGIIYNPNNEEV-----PAL 141
Db 545 EXERWHDILNKEENFQKLSELKGLIILSEKIQKNAEDKNDYNEHOEIVEKLQNAL 604
QY 142 VRARGFDKDALSEGLHKMSLDNQ-----VSILVAKVEEIPKQSVN-----YGD 185
Db 605 IASRWSTQIQEGENTHKITDELQAGQSBILKLEETILSLK-EDVFPQEKLNKLYGD 661

RESULT 6
Q9UAQ7 PRELIMINARY; PRT; 662 AA.
ID Q9UAQ7
AC Q9UAQ7
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-OCT-2001 (T-EMBLrel. 18, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical 75.8 kDa protein.
GN C39F7.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
```

```
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Scheet P., Maggi L., Dubbelde C.;
RT "The sequence of C. elegans cosmid C39F7.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF101310; AAC6214.2; -.
DR WormPep; C39F7.5; CE29087.
DR InterPro; IPR000210; BTB_POZ.
DR SMART; SM00225; BTB; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
KW Hypothetical protein.
SQ SEQUENCE 662 AA; 75799 MW; DD64BBCA0579E102 CRC64;

Query Match 9.5%; Score 93; DB 5; Length 662;
Best Local Similarity 27.5%; Pred. No. 34;
Matches 47; Conservative 30; Mismatches 62; Indels 32; Gaps 9;

QY 1 MLEKSFLLSKQ-----LFLC-----GLGVLMQACTPNTSQNSFLQDVPYWM 46
Db 151 LCEQFLKFEQDAVEIFLKSYYNEQWPNLSEMLEQ-LCTDFGCFKLEEDYKHKYKQ 209
QY 47 NRSEYITQGVDSHIVDGKTEETEK-----IATKRATIRVAQNVHKLKEAYL-S 96
Db 210 D-----LKHGDLIIIEVDGKTEENSESFKRCSSLIASRSKI-IRSLIRKLTEK 264
QY 97 KTNRIKQKTNEM-FIQVTOPIYDSLNVDRGLGIYNPNNEE-VFALVRAR 145
Db 265 STSRPKRIVSELIFFQAFAPFVFSFLYDRLDWSLAPKSEDSISLSQAK 315

RESULT 7
Q87224 PRELIMINARY; PRT; 355 AA.
ID Q87224
AC Q87224
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE LtrC.
OS Lactococcus lactis.
OG Plasmid pMRC01.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1358;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DPC3147;
RX MEDLINE=99000510; PubMed=9767571;
RA Dougherty B.A., Hill C., Weidman J.F., Richardson D.R., Venter J.C.,
RA Ross R.P.;
RT "Sequence and analysis of the 60 kb conjugative, bacteriocin-producing
RT plasmid pMRC01 from Lactococcus lactis DPC3147.";
RL Mol. Microbiol. 29:1029-1038(1998).
DR EMBL; AE001272; AAC56005.1; -.
DR InterPro; IPR006025; ZN_MTPeptidse.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Plasmid.
SQ SEQUENCE 355 AA; 40737 MW; 6CCA54DA50431752 CRC64;

Query Match 9.5%; Score 92.5; DB 2; Length 355;
Best Local Similarity 24.5%; Pred. No. 18;
Matches 38; Conservative 28; Mismatches 50; Indels 39; Gaps 6;

QY 43 WMLQNRSEYITQGVDSHIVDGK--TEETEIKATKRATIRVAQNVHKLKEAYLSKTRN 100
Db 44 YSINNINLIYQNDPDATAVAGFKQWGTDFNRKNKNGEKAIRIAPIKK-----LSEEEK 98
```

QY 101 IKQKITNEMFIOMTQ--PIVD-----SLMNVD-----RLGIYINP 133
 DB 99 IKLTDTDAIVGRIYIPFVDSQSDPLPSARDPVKENLSEVENVDLYKSLKYNQ 158
 QY 134 N-----NEEFALVRARGFQKDALSEGLHKMSLDN 163
 DB 159 NTDIKVSEVLSDVEVKGFRPSTNQIMNESVDN 193

RESULT 8
 Q99968 PRELIMINARY; PRT; 2363 AA.
 AC Q99968;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Tpr.
 GN Tpr.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97177132; PubMed=9024684;
 RA Cordes V.C., Reidenbach S., Rackwitz H.R., Franke W.M.;
 RT "Identification of protein p270/tpr as a constitutive component of the
 RT nuclear pore complex-attached intranuclear filaments."
 RL J. Cell Biol. 136:515-529(1997).
 DR EMBL; U69668; AAB48030.1; -.
 SQ SEQUENCE 2363 AA; 267334 MW; E9BA1C6E78AA35B0 CRC64;

Query Match 9.5%; Score 92.5; DB 4; Length 2363;
 Best Local Similarity 24.1%; Pred. No. 1.7e+02;
 Matches 48; Conservative 38; Mismatches 74; Indels 39; Gaps 10;

QY 17 LGVLMQACTCPNTSQRNSFLQ-----DVPYMLQNRSEYITQGV-----S 58
 DB 1302 LDILPLQANA-ELSEKSGMLQAEKLEEDVKEWKARNQ-HLVSOQKDPDTEYRKLLS 1359
 QY 59 SHIVDGKK-----TEEL-----EKIATKATIRVAQNIHVHKLKEAYLSKTNRIKOKITNEMF 110
 DB 1360 KEVHTKRIQQLTEIGRLKAEIARSNASLTNNQNLQSLKED-LNKVTEKETIQLKDL 1418.
 QY 111 IQM--TQPIYSLMNVDRLGIYINPNNEEVFALVRARGFQKDALSEGLHKMSLDNQAISI 168
 DB 1419 AKIIDIQKVITITQVKIGERYKTYEELKA-----QQDKWETSAAOSSGDHQEQHV 1471
 QY 169 LVAKVBEIFKDSVNGDVK 187
 DB 1472 SVQEMQEL-KETLNQAETK 1489

RESULT 9
 Q8SVU2 PRELIMINARY; PRT; 268 AA.
 AC Q8SVU2;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein ECU04_0860.
 GN ECU04_0860.
 OS Encephalitozoon cuniculi.
 OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
 OX NCBI_TaxID=6035;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1;
 RA Genoscope;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1;
 RX MEDLINE=21574510; PubMed=11719806;
 RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thonarat F.,
 RA Prensier G., Barbe V., Peyretailade E., Brottier P., Wincker P.,
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
 RA Weissenbach J., Vivares C.P.;
 RT "Genome sequence and gene compaction of the eukaryote parasite
 RT Encephalitozoon cuniculi";
 RL Nature 414:450-453(2001).
 DR EMBL; AL590444; CAD25273.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 268 AA; 31477 MW; 574A82F6DDA08705 CRC64;

Query Match 9.4%; Score 91.5; DB 5; Length 268;
 Best Local Similarity 21.6%; Pred. No. 15;
 Matches 35; Conservative 36; Mismatches 56; Indels 35; Gaps 7;

QY 3 EKSFLSKQLFLCGLGLVLMQACTCPNTSQRNSFLQDVPYMLQNRSEYITQGVDSHIV 62
 DB 90 ERFYLESTLVLIKGIALLVDRV--DQEVQRNDFRN-----LEYLYYTDEAIDLTRL 140
 QY 53 DGKKTETIEK---TATKATIRVAQNIHVHKLKE-----AYLSKTNRIKOKITNEMFIQMTQ 115
 DB 141 VGPYTRYVEDPQVILVSRKMPLELNNAVFLNKGLEGVGRSFLADDERLKDVRN----- 191
 QY 116 PIYDSLNMVDRLGIIYNPNNEEVFA-----LVRARGFDKDA 151
 DB 192 -LSEGILOTRIERYV---TEEVYGSIFDLIVKATGMDVDS 229

RESULT 10
 Q9RNA5 PRELIMINARY; PRT; 404 AA.
 AC Q9RNA5;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Uma4.
 GN Uma4.
 OS Microcystis aeruginosa.
 OC Bacteria; Cyanobacteria; Chroococcales; Microcystis.
 OX NCBI_TaxID=1126;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PCC7806;
 RX MEDLINE=20491409; PubMed=11033079;
 RA Tillett D., Dittmann E., Erhard M., von Dohren H., Borner T.,
 RA Neilan B.A.;
 RT "Structural organization of microcystin biosynthesis in Microcystis
 RT aeruginosa PCC7806: an integrated peptide-polyketide synthetase
 RT system";
 RL Chem. Biol. 7:753-764(2000).
 DR EMBL; AF183408; AAF00967.1; -.
 DR InterPro; IPR002560; Transposase.12.
 DR Pfam; PF01610; Transposase.12; 1-
 SQ SEQUENCE 404 AA; 47470 MW; 560506423A8C22B9 CRC64;

Query Match 9.4%; Score 91.5; DB 2; Length 404;
 Best Local Similarity 25.2%; Pred. No. 25;
 Matches 32; Conservative 28; Mismatches 34; Indels 33; Gaps 6;

QY 67 TEEIEKIATKRA-TIRVAQNIHVHKLKEAYLSKTNRIKOKITNEMFIQMTQPIYDSLNMVD 125
 DB 85 SEELDFVAKRTYTKLAENILEQLKEGDILNIS-RNDVTEEEIQRMEDIAEITTD 143
 QY 126 -----RLGI-----YINPNNEEVFALVRAR-----CFDKDASEG 155
 DB 144 LSKLRLGIDETALVKGQKNYCAVLNLDGTGLIALEKRTQELRETLTGWEVL-EQ 202
 QY 156 LHKMSLD 162
 DB 203 IEEVSD 209

```
RESULT 11
Q9NW91 PRELIMINARY; PRT; 450 AA.
ID Q9NW91
AC Q9NW91
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Isogai T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RA "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK001072; BAA91492.1; -
DR InterPro; IPR000299; Band 4.1.
DR PRINTS; PR00373; Band 4.1.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00057; BAND_41_3; 1.
KW Hypothetical protein.
SQ SEQUENCE 450 AA; 51492 MW; 975AFID53C7A7D76 CRC64;

Query Match 9.4%; Score 91.5; DB 4; Length 450;
Best Local Similarity 23.9%; Pred. No. 28;
Matches 48; Conservative 27; Mismatches 65; Indels 65; Gaps 10;

QY 20 LMLQACTCENSRNSFLQDPVYMWLQNRSEYITQGVDSHIVDGKTEETEEIKIAI 75
DB 1 MVVQAAPNRSQR--LLKIPYGLRRSRVERMTEGRGCVHLLDQKLELLVQPKLLA 58
QY 76 KRATIRVAQNIHVHKLKE--AY 94
DB 59 KELLDLVASHFNKLEKEYPGIAFTDGTGHLNQLDRLVLEHDPFKSGPVLVFCVRFY 118
QY 95 LSKTRIKCKITNEM--FIQMTQPIYDSLNVNDRGLGIYINPNNEEVFALV-----RARG-- 146
DB 119 IESISYLNKDNATIELFLLNAKSCIYKELIDVD-----SEVVFELASVILQEAQDP 169
QY 147 FPKDALSEGLHKW-SLDNQAV 166
DB 170 SSNEVVRSDKKLPALPTQAL 190

RESULT 12
Q81LB9 PRELIMINARY; PRT; 8094 AA.
ID Q81LB9
AC Q81LB9
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
```

```
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shalimov S.J., Suh B., Peterson J., Anguoli S.,
RA Ferreira M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RA "Genome sequence of the human malaria parasite Plasmodium
RA falciparum";
RL Nature 419:498-511 (2002).
RL EMBL; AE014821; AAN36939.1; -.
KW Hypothetical protein.
SQ SEQUENCE 8094 AA; 969712 MW; 55822A9C8EE8F151 CRC64;

Query Match 9.4%; Score 91.5; DB 5; Length 8094;
Best Local Similarity 20.1%; Pred. No. 9.3e+02;
Matches 38; Conservative 35; Mismatches 63; Indels 53; Gaps 8;

QY 27 CENTQSRNSFLQDPVYMWLQNRSEYI-----TQGVDSHIVDGKTEETEEIKIA 74
DB 5434 CERLQEDNNVEDM-----NTKEHIECNCTNEEVLEKLNKIDNSSVIESK-EKINEVL 5486
QY 75 TKRATIRVAQNIHVHKLKEAVLSKTRIK-----OKITNEMFIQMTQPI 116
DB 5467 FNKIIYKLVKLFKRLKEKIEDKNNIKNYNCTSEYIKQTYINRIQKNHLEF-----EM 5542
QY 117 IYDSLNVNDRGLGIYINPNNEEVFALVARGFDKDALSEGLHKMSLDNQAVSILVAKVEEI 176
DB 5543 LIFNINNKCKNSFISNNKNE-----NPD---LNINLWLNKANEENKFL-IFTLNILNY 5590
QY 177 FKDSVNYGD 185
DB 5591 ENDFNYND 5599

RESULT 13
Q81BS0 PRELIMINARY; PRT; 5779 AA.
ID Q81BS0
AC Q81BS0
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN MAL7P1.89.
OC Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844506; CAD50928.1; -.
KW Hypothetical protein.
SQ SEQUENCE 5779 AA; 690899 MW; F930AC1A8C7B6E4C CRC64;

Query Match 9.3%; Score 91; DB 5; Length 5779;
Best Local Similarity 21.5%; Pred. No. 6.8e+02;
Matches 48; Conservative 39; Mismatches 74; Indels 62; Gaps 8;

QY 1 MLEKSFLLSKQFLCGLGLVLMQLQACTCENSRNSFLQDPVYMWLQNRSEYI----- 51
DB 3934 MLSKKYLNK-----GINV-----CINSKYLHFLERFDYLFYFKKKEYDKHDIYSK 3981
QY 52 -----ITQGVDSHIVDGKTEETEEIKIAIKRATIRVAQNIHVHKL 90
DB 3982 ALHLKHQCEBDIKMNKLNLIQIILNSTNEMKKRVNEIER-DTKDAYIKQSE-----I 4035
QY 91 KEAYLSKTRNIK-----QKITNEMFIQMTQPIYDSLNVNDRGLGI-----YINPNNEE 137
DB 4036 KKKENDVKNKIKITNLKNVNVNEISKSFLLNDSLNKLNKLVKVEHLRELKAFVNPSPV 4095
QY 138 VFALVARGPDKDALSEGLHKMSLDNQAVSILVAKVEEIPKDS 180
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:28:11, Search time 26.0105 Seconds
(without alignments)
878.747 Million cell updates/sec

Title: US-10-080-113-4

Perfect score: 711

Sequence: 1 MNISVNPYMAVFWVFVLL.....AQMPVFDELKRVAMGLGS 144

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	711	100.0	144	19	AAW98353 H. pylori GHPO 670
2	685	96.3	144	18	AAW55448 H. pylori ORF 019p
3	585	82.3	123	18	AAW55208 H. pylori ORF 019p
4	146	20.5	148	21	AAG09710 Arabidopsis thalia
5	146	20.5	148	21	AAW25565 Arabidopsis thalia
6	146	20.5	148	21	AAW42453 Arabidopsis thalia
7	146	20.5	213	21	AAG09709 Arabidopsis thalia
8	146	20.5	213	21	AAW25565 Arabidopsis thalia
9	146	20.5	213	21	AAW42452 Arabidopsis thalia

10	146	20.5	219	21	AAG09708 Arabidopsis thalia
11	146	20.5	219	21	AAW25564 Arabidopsis thalia
12	146	20.5	219	21	AAW42451 Arabidopsis thalia
13	125	17.6	156	22	AAU38475 Salmonella typhi c
14	123	17.3	156	22	AAU34809 E. coli cellular p
15	123	17.3	156	22	AAW98292 Escherichia coli p
16	120.5	16.9	156	22	Pseudomonas aerugi
17	108	15.2	660	22	Peptide #3468 enco
18	108	15.2	660	23	Human peptide enco
19	107.5	15.1	178	23	Staphylococcus epi
20	106	14.9	1177	22	Putative P. abyssal
21	105.5	14.8	156	24	N. gonorrhoeae am
22	102	14.3	1001	23	Aspergillus fumiga
23	101	14.2	170	23	Listeria monocytog
24	101	14.2	1690	22	Drosophila melanog
25	101	14.2	1690	22	Drosophila melanog
26	99.5	14.0	162	22	Novel human diagno
27	98.5	13.9	226	23	Human macroprotein
28	98.5	13.9	470	24	Human structural a
29	98.5	13.9	945	20	Human secreted pro
30	98.5	13.9	945	20	Human polypeptide
31	98.5	13.9	1711	23	CDC42-binding prot
32	98.5	13.9	1728	21	Human ORFX ORF183
33	98.5	13.9	1797	22	Novel human diagno
34	98	13.8	525	23	Bifidobacterium lo
35	98	13.8	563	23	Staphylococcus epi
36	97.5	13.7	188	22	C glutamic prote
37	97	13.6	409	24	Human DITP intrac
38	97	13.6	1719	23	Human PKIN-2 prote
39	97	13.6	1770	23	Human kinase and p
40	95.5	13.4	168	22	Staphylococcus aur
41	95.5	13.4	173	22	Staphylococcus aur
42	95.5	13.4	173	22	Staphylococcus aur
43	95.5	13.4	173	22	Staphylococcus aur
44	95.5	13.4	1694	23	Human GDMF-1 orth
45	95	13.4	1881	23	Candida albicans e

ALIGNMENTS

RESULT 1

AAW98353
ID AAW98353 standard; Protein; 144 AA.

AC AAW98353;

XX 31-MAR-1999 (first entry)

DT H. pylori GHPO 670 protein.

DE GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;

XX peptic ulcer disease.

OS Helicobacter pylori.

XX WO9843478-A1.

XX 08-OCT-1998.

XX 01-APR-1998; 98WO-US06371.

XX 29-JUL-1997; 97US-0902615.

XX 01-APR-1997; 97US-0833457.

XX 24-JUN-1997; 97US-0881227.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

XX Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;

XX WPI; 1998-542293/46.

XX N-PSDB; AA14072.

XX New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases
 XX
 PS Claim 8; Page 508-509; 2054pp; English.
 XX This sequence represents a Helicobacter pylori GHPO protein of the
 CC invention. The polypeptides can be used for preventing or treating
 CC Helicobacter infections, and gastroduodenal diseases associated with
 CC these infections, including acute, chronic, and atrophic gastritis, and
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 CC used for the production of antibodies. The products can also be used for
 CC detection and diagnosis.
 XX
 SQ Sequence 144 AA;
 Query Match 100.0%; Score 711; DB 19; Length 144;
 Best Local Similarity 100.0%; Pred. No. 3.8e-59;
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MNISVNPYLMVAVFVVFVLLWMNVVYRPLLAFLAFMDNRQAEIKDSLAKIKTDNAQSVEI 60
 Db 1 MNISVNPYLMVAVFVVFVLLWMNVVYRPLLAFLAFMDNRQAEIKDSLAKIKTDNAQSVEI 60
 Qy 61 GHOIEALLKEAAEKREIIAEAIQKATESYDAVIKQENELNQEFAFAKOLQNEKQALK 120
 Db 61 GHOIEALLKEAAEKREIIAEAIQKATESYDAVIKQENELNQEFAFAKOLQNEKQALK 120
 Qy 121 EQLQAQMPVFEDELNKRVMGLGS 144
 Db 121 EQLQAQMPVFEDELNKRVMGLGS 144
 RESULT 2
 AAW55448
 ID AAW55448 standard; Protein; 144 AA.
 AC AAW55448;
 XX
 DT 24-JUN-1998 (first entry)
 DE H. pylori ORF 01gp11016_4103403_c2_13 secreted protein.
 XX
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 XX
 OS Helicobacter pylori.
 XX
 PN WO9737044-A1.
 XX
 PD 09-OCT-1997.
 XX
 PF 27-MAR-1997; 97WO-US05223.
 XX
 PR 06-DEC-1996; 96US-0761318.
 PR 29-MAR-1996; 96US-0625811.
 PR 02-APR-1996; 96US-0758731.
 PR 25-OCT-1996; 96US-0736905.
 PR 28-OCT-1996; 96US-0738859.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Alm RA, Smith D;
 XX
 DR WPI; 1997-503122/46.
 DR N-PSDB; AAV24857.
 XX
 XX Helicobacter pylori nucleic acid sequences and encoded
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
 PT infection and for diagnosis of H. pylori infection
 XX

PS Claims 14,94; Page 655-656; 1145pp; English.
 XX This sequence is a H. pylori secreted protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds. The
 CC useful as potential H. pylori life cycle activators or inhibitors. The
 CC DNA and probes derived from it may be used for the identification of
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences and
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of H. pylori-specific antigens. The genomic sequence of
 CC H. pylori (ATCC 35679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions
 CC defined by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or expected
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from H. pylori by PCR
 CC amplification for recombinant polypeptide production, e.g. in E. coli
 CC hosts.
 XX
 SQ Sequence 144 AA;
 Query Match 96.3%; Score 685; DB 18; Length 144;
 Best Local Similarity 95.8%; Pred. No. 1.1e-56;
 Matches 138; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 MNISVNPYLMVAVFVVFVLLWMNVVYRPLLAFLAFMDNRQAEIKDSLAKIKTDNAQSVEI 60
 Db 1 MNISVNPYLMVAVFVVFVLLWMNVVYRPLLAFLAFMDNRQAEIKDSLAKIKTDNTOSVEI 60
 Qy 61 GHOIEALLKEAAEKREIIAEAIQKATESYDAVIKQENELNQEFAFAKOLQNEKQALK 120
 Db 61 GHOIEALLKEAAEKREIIAEAIQKATESYDAVIKQENELNQEFAFAKOLQNEKQILK 120
 Qy 121 EQLQAQMPVFEDELNKRVMGLGS 144
 Db 121 EQLQAQMPVFEDELNKRVMGLGS 144
 RESULT 3
 AAW55208
 ID AAW55208 standard; Protein; 123 AA.
 XX
 AC AAW55208;
 XX
 DT 15-JUN-1998 (first entry)
 DE H. pylori ORF 01gp11016orf13 protein.
 XX
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 XX
 OS Helicobacter pylori.
 XX
 PN WO9737044-A1.
 XX
 PD 09-OCT-1997.
 XX
 PF 27-MAR-1997; 97WO-US05223.
 XX
 PR 06-DEC-1996; 96US-0761318.
 PR 29-MAR-1996; 96US-0625811.
 PR 02-APR-1996; 96US-0758731.
 PR 25-OCT-1996; 96US-0736905.
 PR 28-OCT-1996; 96US-0738859.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Alm RA, Smith D;
 XX

```

XX WPI; 1997-503122/46.
DR N-PSDB; AAV24617.
XX
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX
XX Claim 14; Page 458-459; 1145pp; English.
XX
CC This sequence is a H. pylori protein of unspecified function.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
XX
XX Sequence 123 AA;
XX
Query Match 82.3%; Score 585; DB 18; Length 123;
Best Local Similarity 95.1%; Pred. No. 2.1e-47;
Matches 117; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 22 WAMVWVYRPLAFMNDQAEIKDSIAKIKTDAQSVSEIGHQIEALLKEAAEKREITAE 81
Db |||||
1 WAMVWVYRPLAFMNDQAEIKDSIAKIKTNTQSVSEIGHQIEALLKEAAEKREMLAE 60
QY 82 ATOKATESYDAVIKQKENELNQEFAFAKQLONEKQALKEQLQAOVPFEDLNKRVAMG 141
Db |||||
61 A:OKATESYDAVIKQKENELNQEFAFAKQLONEKQILKEQLQAOQMTVFEDLNKRVAMG 120
QY 142 LGS 144
Db |||
121 LGS 123

RESULT 4
AAG09710
ID AAG09710 standard; Protein; 148 AA.
XX
AC AAG09710;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 7747.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.

```

```

PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143342.

```

PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144003.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144085.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144334.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159323.
PR	19-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144633.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144634.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160814.
PR	23-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160981.
PR	26-JUL-1999;	99US-0145276.	PR	22-OCT-1999;	99US-0160981.
PR	27-JUL-1999;	99US-0145913.	PR	22-OCT-1999;	99US-0160989.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145919.	PR	25-OCT-1999;	99US-0161405.
PR	28-JUL-1999;	99US-0145951.	PR	25-OCT-1999;	99US-0161406.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161359.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146389.	PR	26-OCT-1999;	99US-0161361.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161920.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161992.
PR	04-AUG-1999;	99US-0147302.	PR	28-OCT-1999;	99US-0161993.
PR	05-AUG-1999;	99US-0147192.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148584.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149802.			
PR	23-AUG-1999;	99US-0149920.			
PR	25-AUG-1999;	99US-0150366.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	30-AUG-1999;	99US-0151080.			
PR	30-AUG-1999;	99US-0151303.			
PR	31-AUG-1999;	99US-0151438.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
PR	13-SEP-1999;	99US-0153758.			
PR	15-SEP-1999;	99US-0154018.			

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135639.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0139750.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141844.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

```

PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 20.5%; Score 146; DB 21; Length 148;
Best Local Similarity 25.6%; Pred. No. 4.5e-06;
Matches 34; Conservative 36; Mismatches 55; Indels 8; Gaps 2;

QY 10 MAVFVYVLLWANNVYRLLAFMDNRQAEINKDSLAKIKTDNAQSVEIGHQIEALLK 69
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17 LPIIVVEFLFMALDKVYVPLGNFMQORDASIKELASVKDTSTEVKELDEQAAAVMR 76
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 70 EAAKKREIIAIAQKATESYDAVIKQK----ENELNQEFEPAPAKQLCNKQALKEOLOA 125
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 77 AA----RAETAAALNNKKTQTOVEEKLAEGRKKVBEELKEALASLESQKEETIKALDS 132
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 QMFVFEDELNKRK 138
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 133 QTAALSEDIVKKV 145

RESULT 6
AAG42453
ID AAG42453 standard; Protein; 148 AA.
XX
AC AAG42453;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 52947.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX

```

```

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130448.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139432.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.

```


PF	25-FEB-2000;	2000EP-0301439.	
XX	25-FEB-1999;	99US-0121825.	99US-0142920.
PR	05-MAR-1999;	99US-0123180.	99US-0142977.
PR	09-MAR-1999;	99US-0123548.	99US-0143542.
PR	23-MAR-1999;	99US-0125788.	99US-0143624.
PR	25-MAR-1999;	99US-0126264.	99US-0144005.
PR	29-MAR-1999;	99US-0126785.	99US-0144085.
PR	01-APR-1999;	99US-0127462.	99US-0144086.
PR	06-APR-1999;	99US-0128234.	99US-0144325.
PR	08-APR-1999;	99US-0128714.	99US-0144332.
PR	16-APR-1999;	99US-0129845.	99US-0144333.
PR	19-APR-1999;	99US-0130077.	99US-0144334.
PR	21-APR-1999;	99US-0130449.	99US-0144335.
PR	23-APR-1999;	99US-0130510.	99US-0144332.
PR	28-APR-1999;	99US-0130891.	99US-0144632.
PR	30-APR-1999;	99US-0132048.	99US-0144884.
PR	04-MAY-1999;	99US-0132484.	99US-0144814.
PR	05-MAY-1999;	99US-0132485.	99US-0145086.
PR	06-MAY-1999;	99US-0132486.	99US-0145088.
PR	07-MAY-1999;	99US-0132487.	99US-0145085.
PR	11-MAY-1999;	99US-0134256.	99US-0145087.
PR	14-MAY-1999;	99US-0134218.	99US-0145089.
PR	14-MAY-1999;	99US-0134219.	99US-0145192.
PR	14-MAY-1999;	99US-0134221.	99US-0145145.
PR	18-MAY-1999;	99US-0134370.	99US-0145218.
PR	19-MAY-1999;	99US-0134768.	99US-0145224.
PR	20-MAY-1999;	99US-0134941.	99US-0145276.
PR	20-MAY-1999;	99US-0135124.	99US-0145913.
PR	21-MAY-1999;	99US-0135353.	99US-0145918.
PR	24-MAY-1999;	99US-0135629.	99US-0145919.
PR	25-MAY-1999;	99US-0136021.	99US-0145951.
PR	27-MAY-1999;	99US-0136392.	99US-0146386.
PR	28-MAY-1999;	99US-0136782.	99US-0146388.
PR	01-JUN-1999;	99US-0137222.	99US-0146389.
PR	03-JUN-1999;	99US-0137528.	99US-0147038.
PR	04-JUN-1999;	99US-0137502.	99US-0147204.
PR	07-JUN-1999;	99US-0137724.	99US-0147302.
PR	08-JUN-1999;	99US-0138094.	99US-0147192.
PR	10-JUN-1999;	99US-0138540.	99US-0147260.
PR	10-JUN-1999;	99US-0138847.	99US-0147303.
PR	14-JUN-1999;	99US-0139119.	99US-0147416.
PR	16-JUN-1999;	99US-0139452.	99US-0147493.
PR	16-JUN-1999;	99US-0139453.	99US-0147935.
PR	17-JUN-1999;	99US-0139492.	99US-0148171.
PR	18-JUN-1999;	99US-0139454.	99US-0148319.
PR	18-JUN-1999;	99US-0139455.	99US-0148341.
PR	18-JUN-1999;	99US-0139456.	99US-0148565.
PR	18-JUN-1999;	99US-0139457.	99US-0148684.
PR	18-JUN-1999;	99US-0139458.	99US-0149175.
PR	18-JUN-1999;	99US-0139459.	99US-0149175.
PR	18-JUN-1999;	99US-0139460.	99US-0149426.
PR	18-JUN-1999;	99US-0139461.	99US-0149722.
PR	18-JUN-1999;	99US-0139462.	99US-0149722.
PR	18-JUN-1999;	99US-0139463.	99US-0149929.
PR	18-JUN-1999;	99US-0139750.	99US-0149902.
PR	18-JUN-1999;	99US-0139763.	99US-0149930.
PR	21-JUN-1999;	99US-0139817.	99US-0150566.
PR	21-JUN-1999;	99US-0139819.	99US-0150884.
PR	23-JUN-1999;	99US-0140353.	99US-0151065.
PR	23-JUN-1999;	99US-0140354.	99US-0151065.
PR	24-JUN-1999;	99US-0140695.	99US-0151066.
PR	28-JUN-1999;	99US-0140823.	99US-0151080.
PR	29-JUN-1999;	99US-0140991.	99US-0151303.
PR	30-JUN-1999;	99US-0141287.	99US-0151438.
PR	01-JUL-1999;	99US-0141842.	99US-0151930.
PR	01-JUL-1999;	99US-0142154.	99US-0152363.
PR	02-JUL-1999;	99US-0142055.	99US-0153070.
PR	06-JUL-1999;	99US-0142390.	99US-0153758.
PR	08-JUL-1999;	99US-0142803.	99US-0154018.
PR	08-JUL-1999;	99US-0142803.	99US-0154039.
PR	08-JUL-1999;	99US-0142803.	99US-0154779.
PR	08-JUL-1999;	99US-0142803.	99US-0155139.
PR	08-JUL-1999;	99US-0142803.	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	99US-0155659.

PR 08-JUL-1999;	99US-0142803.	PR 24-SEP-1999;	99US-0155659.
PR 09-JUL-1999;	99US-0142920.	PR 26-SEP-1999;	99US-0156458.
PR 12-JUL-1999;	99US-0142977.	PR 29-SEP-1999;	99US-0156596.
PR 13-JUL-1999;	99US-0143342.	PR 04-OCT-1999;	99US-0157117.
PR 14-JUL-1999;	99US-0143624.	PR 05-OCT-1999;	99US-0157753.
PR 15-JUL-1999;	99US-0144005.	PR 06-OCT-1999;	99US-0157865.
PR 16-JUL-1999;	99US-0144085.	PR 07-OCT-1999;	99US-0158029.
PR 16-JUL-1999;	99US-0144086.	PR 08-OCT-1999;	99US-0158232.
PR 19-JUL-1999;	99US-0144325.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159329.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159637.
PR 20-JUL-1999;	99US-0144684.	PR 14-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0144814.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0160741.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160767.
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160814.
PR 23-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160981.
PR 28-JUL-1999;	99US-0145276.	PR 22-OCT-1999;	99US-0160989.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161404.
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161405.
PR 28-JUL-1999;	99US-0145919.	PR 25-OCT-1999;	99US-0161406.
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161359.
PR 02-AUG-1999;	99US-0146386.	PR 26-OCT-1999;	99US-0161360.
PR 02-AUG-1999;	99US-0146388.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0146389.	PR 26-OCT-1999;	99US-0161920.
PR 03-AUG-1999;	99US-0147038.	PR 28-OCT-1999;	99US-0161992.
PR 04-AUG-1999;	99US-0147204.	PR 28-OCT-1999;	99US-0161993.
PR 04-AUG-1999;	99US-0147302.	PR 29-OCT-1999;	99US-0162142.
PR 05-AUG-1999;	99US-0147192.		
PR 05-AUG-1999;	99US-0147260.		
PR 06-AUG-1999;	99US-0147303.		
PR 06-AUG-1999;	99US-0147416.		
PR 08-AUG-1999;	99US-0147493.		
PR 09-AUG-1999;	99US-0147935.		
PR 10-AUG-1999;	99US-0148171.		
PR 11-AUG-1999;	99US-0148319.		
PR 12-AUG-1999;	99US-0148341.		
PR 13-AUG-1999;	99US-0148565.		
PR 13-AUG-1999;	99US-0148684.		
PR 16-AUG-1999;	99US-0149368.		
PR 17-AUG-1999;	99US-0149175.		
PR 18-AUG-1999;	99US-0149426.		
PR 20-AUG-1999;	99US-0149722.		
PR 20-AUG-1999;	99US-0149723.		
PR 20-AUG-1999;	99US-0149929.		
PR 23-AUG-1999;	99US-0149902.		
PR 23-AUG-1999;	99US-0149930.		
PR 25-AUG-1999;	99US-0150566.		
PR 26-AUG-1999;	99US-0150884.		
PR 27-AUG-1999;	99US-0151065.		
PR 27-AUG-1999;	99US-0151066.		
PR 27-AUG-1999;	99US-0151080.		
PR 30-AUG-1999;	99US-0151303.		
PR 31-AUG-1999;	99US-0151438.		
PR 01-SEP-1999;	99US-0151930.		
PR 07-SEP-1999;	99US-0152363.		
PR 10-SEP-1999;	99US-0153070.		
PR 13-SEP-1999;	99US-0153758.		
PR 15-SEP-1999;	99US-0154018.		
PR 16-SEP-1999;	99US-0154039.		
PR 20-SEP-1999;	99US-0154779.		
PR 22-SEP-1999;	99US-0155139.		
PR 23-SEP-1999;	99US-0155486.		

Query Match 20.5%; Score 146; DB 21; Length 213;

Best Local Similarity 25.6%; Pred. No. 6.9e-06;

Matches 34; Conservative 36; Mismatches 55; Indels 8; Gaps 2;

QY 10 MAVFVFPVLLMNMVVRPLAFMDNRQASIKSLAKIKTDNACSVEIGHQIEALIK 69

Db 82 LPITVVEFLFMFALDKVYSPGLGNFMDQDASIKELASVKDTSTEVKELDEQAAAVR 141

QY 70 EAAEKRRRIIAEAIQKATSYDAVKKK---ENELNOEFAFAKQIQNEKQAKQLOA 125

Db 142 AA---RAETAAALNKKMKKETOVEEKLAEGRKKVEELKEALASLESQKETIKALDS 197

QY 126 QMPVFEDELNKR 138

Db 198 QIAALSEDIKKV 210

RESULT 9

AG42452

ID AAG42452 standard; Protein; 213 AA.

XX AAG42452;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 52946.

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
FF 25-FEB-1999; 99US-0121825.
FF 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149368.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152163.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.

us-10-080-113-4.rag

Fri Dec 12 13:23:15 2003

```

PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 20.5%; Score 146; DB 21; Length 213;
Best Local Similarity 25.6%; Pred. No. 6.9e-06;
Matches 34; Conservative 36; Mismatches 55; Indels 8; Gaps 2;

QY 10 MAVVVFVLLMANNVWVYRLLAFMDNRQAEIKDSIAKIKTDNAOSVEIGHQIEALLK 69
DB 82 LPIIVVEFLFMFALDKVYISPLGNFMDQDASIKELASVKDSTEVKELDEQAAMVR 141
QY 70 EAAEKRRRIIAEAIQKATESYDAVIKQK----ENELNQEFFAFAKQIQNEKQALKEQLQA 125
DB 142 AA----RAEIAALNKKMKETQVEVEEKLAEGRKKVBEELKEALASLESQKEETIKALDS 197
QY 126 QMPVEFDELNKKV 138
DB 198 QIAALSEDIVKKV 210

RESULT 10
AAG09708
ID AAG09708 standard; Protein; 219 AA.
XX AC AAG09708;
XX DT
XX 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 7745.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX EP1033405-R2.

```

```

XX PD
XX PF
XX XX
PR 06-SEP-2000.
PR 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135253.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 03-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.

```


PN EP1033405-A2.
XX
XX
PD
XX
XX
PF 25-FEB-2000; 2000EP-030439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 16-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149388.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149920.
PR 26-AUG-1999; 99US-0150586.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.

DE Arabidopsis thaliana protein fragment SEQ ID NO: 52945.
XX
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
OS Arabidopsis thaliana.

PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	02-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	08-JUL-1999;	99US-0142803.	PR	28-SEP-1999;	99US-0156458.
PR	09-JUL-1999;	99US-0142920.	PR	29-SEP-1999;	99US-0156596.
PR	12-JUL-1999;	99US-0142977.	PR	04-OCT-1999;	99US-0157117.
PR	13-JUL-1999;	99US-0143542.	PR	05-OCT-1999;	99US-0157753.
PR	14-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160744.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	28-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	27-AUG-1999;	99US-0151080.			
PR	27-AUG-1999;	99US-0151080.			
PR	30-AUG-1999;	99US-0151303.			
PR	31-AUG-1999;	99US-0151438.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
PR	13-SEP-1999;	99US-0153758.			
PR	15-SEP-1999;	99US-0154018.			
PR	16-SEP-1999;	99US-0154039.			
PR	20-SEP-1999;	99US-0154779.			
PR	22-SEP-1999;	99US-0155139.			
PR	23-SEP-1999;	99US-0155486.			
PR	24-SEP-1999;	99US-0155659.			
PR	28-SEP-1999;	99US-0156458.			
PR	29-SEP-1999;	99US-0156596.			
PR	04-OCT-1999;	99US-0157117.			
PR	05-OCT-1999;	99US-0157753.			
PR	06-OCT-1999;	99US-0157865.			
PR	07-OCT-1999;	99US-0158029.			
PR	08-OCT-1999;	99US-0158232.			
PR	12-OCT-1999;	99US-0158369.			
PR	13-OCT-1999;	99US-0159293.			
PR	13-OCT-1999;	99US-0159294.			
PR	13-OCT-1999;	99US-0159295.			
PR	14-OCT-1999;	99US-0159329.			
PR	14-OCT-1999;	99US-0159330.			
PR	14-OCT-1999;	99US-0159331.			
PR	14-OCT-1999;	99US-0159637.			
PR	14-OCT-1999;	99US-0159638.			
PR	18-OCT-1999;	99US-0159584.			
PR	21-OCT-1999;	99US-0160744.			
PR	21-OCT-1999;	99US-0160767.			
PR	21-OCT-1999;	99US-0160768.			
PR	21-OCT-1999;	99US-0160770.			
PR	21-OCT-1999;	99US-0160814.			
PR	21-OCT-1999;	99US-0160815.			
PR	22-OCT-1999;	99US-0160980.			
PR	22-OCT-1999;	99US-0160981.			
PR	22-OCT-1999;	99US-0160989.			
PR	25-OCT-1999;	99US-0161404.			
PR	25-OCT-1999;	99US-0161405.			
PR	25-OCT-1999;	99US-0161406.			
PR	26-OCT-1999;	99US-0161359.			
PR	26-OCT-1999;	99US-0161360.			
PR	26-OCT-1999;	99US-0161361.			
PR	28-OCT-1999;	99US-0161920.			
PR	28-OCT-1999;	99US-0161992.			
PR	28-OCT-1999;	99US-0161993.			
PR	29-OCT-1999;	99US-0162142.			
Query Match 20.5%; Score 146; DB 21; Length 219;					
Best Local Similarity 25.6%; Pred. No. 7.1e-06;					
Matches 34; Conservative 36; Mismatches 55; Indels 8; Gaps 2;					
QY	10	MAVVFVFLILWAMNVMVYRPLAFMDNRQAEIKDSIAKIKTDNAQSVFICHOIEALLK 69			
Db	88	LPFIIVVEFLFMFALDKVYVYPLGNFMFDORDASIKELASVKDTSTEVKDELDEQAAAVMR 147			
QY	70	EAAEKREITIAEIAOKATESYDAVIKQK-----ENELNOEFAFAKOLQNEKALKEQLOA 125			
Db	148	AA----RAETAAALNKKMKETQVEVEEXLAEGRKVVEELKEALASLESQKEETIKALDS 203			
QY	126	QMPVFEDELNKEV 138			
Db	204	QIAALSEDIVKKV 216			
RESULT 13					
AAU38475					
ID	AAU38475 standard; Protein; 156 AA.				
XX	AAU38475;				
AC	AAU38475;				
XX	14-FEB-2002 (first entry)				
DT	Salmonella typhi cellular proliferation protein #366.				
XX	Antisense; prokaryotic cellular proliferation protein;				
KW	antibiotic; antibacterial; drug design.				
XX	Salmonella typhi.				
OS	Salmonella typhi.				

XX WO200170955-A2.
 XX
 XX
 PD 27-SEP-2001.
 XX
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 XX 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 PI
 XX WPI; 2001-611495/70.
 DR N-PSDB; AAS56334.
 DR
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 PT
 PS Example 3; Seq ID No 14068; 51pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: the sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 156 AA;
 SQ
 Query Match 17.6%; Score 125; DB 22; Length 156;
 Best Local Similarity 28.5%; Pred. No. 0.00044;
 Matches 41; Conservative 30; Mismatches 51; Indels 22; Gaps 5;
 QY 1 MNISVNPVLMVVFVVFVLLWMNVVYRPLLAFLMNDNRQAEIKDSLAKIKTDNAOSVEI 60
 DB 1 MNLNATIXGQAIATILFV--WFCMKYVWPLMAIEKQKEIADGLA-----SAER 49
 QY 61 GHQIEALLK-EAAEKREIIIAE---IQKATESYDAVIKQKENELNQEFAFAKQNEK 116
 DB 50 AHKLDLAKASATDQ-LKAKAEQVITIEQANKRBAQILDEAKTEAEQERTKIVAQAQAEI 109
 QY 117 QALKEQLQAQMEVFVEDELNKRVM 140
 DB 110 EAERKRAR-----EELRKQVAI 126
 RESULT 14
 AAU34809
 ID AAU34809 standard; Protein; 156 AA.
 XX
 AC AAU34809;
 XX

DT 14-FEB-2002 (first entry)
 XX
 XX E. coli cellular proliferation protein #390.
 XX
 XX Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 KW
 XX Escherichia coli.
 OS
 XX WO200170955-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX 21-MAR-2001; 2001WO-US09180.
 PF
 XX 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 PA
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 PI
 XX WPI; 2001-611495/70.
 DR N-PSDB; AAS52668.
 DR
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 PT
 PS Example 3; Seq ID No 10402; 51pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: the sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 156 AA;
 SQ
 Query Match 17.3%; Score 123; DB 22; Length 156;
 Best Local Similarity 27.1%; Pred. No. 0.00068;
 Matches 39; Conservative 27; Mismatches 56; Indels 22; Gaps 4;
 QY 1 MNISVNPVLMVVFVVFVLLWMNVVYRPLLAFLMNDNRQAEIKDSLAKIKTDNAOSVEI 60
 DB 1 MNLNATILGQAIATILFVLCMK---YVWPLMAIEKQKEIADGLA-----SAER 49
 QY 61 GHQIEALLK-EAAEKREIIIAE---IQKATESYDAVIKQKENELNQEFAFAKQNEK 120
 DB 50 AHKLDLAKASA-----TDQLKAKAEQVITIEQANKRBAQILDEAKAEAEQERTKIV 102
 QY 121 EQLQAQMEV---FEDELNKRVM 140
 DB 103 AQAQAEIEAEKREAEELRKQVAI 126

RESULT 15
AAG98292
ID AAG98292 standard; Protein; 156 AA.

XX AC AAG98292;
XX 21-SEP-2001 (first entry)
XX Escherichia coli protein sequence SEQ ID NO:340.
XX Escherichia coli; identification; proliferation; microorganism;
XX antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
XX bacterial growth inhibition.

XX OS Escherichia coli.

XX WO200148209-A2.

XX PD 05-JUL-2001.

XX PF 19-DEC-2000; 2000WO-US34419.

XX PR 23-DEC-1999; 99US-0173005.

XX PA (ELIT-) ELITFA PHARM INC.

XX FI Forsyth RA, Ohlsen KL, Zyskind JW;

XX DR WPI; 2001-457376/49.

XX DR N-PSDB; AAH81348.

XX PT Novel nucleic acids encoding proteins required for Escherichia coli
XX proliferation, useful for screening for antimicrobial agents -

XX FS Claim 19; Page 470-471; 596pp; English.

XX CC The present invention describes a purified or isolated nucleic acid
XX sequence (I) consisting essentially of one of the 93 nucleotide sequences
XX given in AAH81202 to AAH81294, where expression of the nucleic acid in a
XX microorganism is capable of inhibiting proliferation of a microorganism.
XX (I) have antibacterial and antibiotic activities, and can be used in
XX gene therapy. Expression of (I) in a microorganism inhibits proliferation
XX of the microorganism, and the manufactured antibiotic is useful for
XX reducing the activity or level of a gene product required for
XX proliferation of a microorganism in a subject, specifically humans. The
XX nucleic acids that inhibit bacterial growth or proliferation can be used
XX as antisense therapeutics for killing bacteria. In addition to
XX therapeutic applications, the nucleic acid sequences complementary to
XX sequences required for proliferation can be used as diagnostic tools.
XX For example, nucleic acid probes complementary to proliferation-required
XX sequences that are specific for particular species of microorganisms can
XX be used as probes to identify particular microorganism species in
XX clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
XX proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491
XX represent oligonucleotides, which are used in the exemplification of the
XX present invention.

XX SQ Sequence 156 AA;

Query Match 17.3%; Score 123; DB 22; Length 156;
Best Local Similarity 27.1%; Pred. No. 0.00068;
Matches 39; Conservative 27; Mismatches 56; Indels 22; Gaps 4;

QY 1 MNISVNYLVAVVVFVLLWAMVWYRPLAFMDNFAEIKDSLAKIKTDNAQSVEI 60

DB 1 MNLNATILGQALAFVLFVFCMK---YVWPLMAIEKEQXELADGLA-----SAER 49

QY 61 GHQTEALLKKAERKRRIIAEAIQKATESYDAVIOKXENELNQEFPAKQLONEKQALK 120

DB 50 AHKDLDLAKASA-----TDQLKKAFAQVILIQANKRRSQILDEAKAEQERTKIV 102

QY 121 ECLQAPMPV---FEDELNKRKVM 140
DB 103 AQAQAEIEAEKRRARESLRKQVAI 126

Search completed: December 10, 2003, 18:33:46
Job time : 27.0105 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:31:22 ; Search time 9.04712 Seconds
(without alignments)
673.448 Million cell updates/sec

Title: US-10-080-113-4

Perfect score: 711

Sequence: 1 MNISVNPYLMAVVFVVFVLL.....AQMVFDELNKRVAMGLS 144

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119.5	16.8	165	4	US-09-252-991A-19000
2	108	15.2	3878	4	US-09-914-259-11
3	107.5	15.1	178	4	US-09-134-001C-4229
4	103.5	14.6	159	4	US-09-328-352-4908
5	98	13.8	563	4	US-09-134-001C-2916
6	94	13.2	961	4	US-09-914-259-66
7	92	12.9	976	3	US-09-104-324B-4
8	91	12.8	208	4	US-09-328-352-5515
9	91	12.8	873	4	US-09-336-447A-13
10	90.5	12.7	608	2	US-08-736-770-1
11	90.5	12.7	900	2	US-08-630-822A-62
12	90.5	12.7	900	2	US-09-005-069-62
13	90.5	12.7	900	4	US-09-171-156A-21
14	90.5	12.7	900	4	US-09-004-730A-21
15	90.5	12.7	900	4	US-08-981-759A-21
16	89.5	12.6	1044	4	US-09-107-532A-5229
17	89	12.5	476	4	US-09-252-991A-24392
18	89	12.5	703	4	US-09-252-991A-17865
19	89	12.5	1939	4	US-09-310-187A-1
20	88.5	12.4	124	2	US-08-743-200-12
21	88.5	12.4	1184	3	US-09-541-782-2
22	88.5	12.4	1184	4	US-09-723-820-2
23	88	12.4	361	4	US-09-328-352-4225
24	88	12.4	466	4	US-09-610-401-3
25	88	12.4	466	4	US-09-167-206-12
26	88	12.4	1886	4	US-08-938-105-3
27	87.5	12.3	831	4	US-09-336-447A-1

28 87.5 12.3 2101 1 US-08-466-390-4 Sequence 4, Appli
29 87.5 12.3 2101 1 US-08-470-950-4 Sequence 4, Appli
30 87.5 12.3 2101 1 US-08-467-781-4 Sequence 4, Appli
31 87.5 12.3 2101 1 US-08-195-487-4 Sequence 4, Appli
32 87.5 12.3 2101 2 US-08-483-924-4 Sequence 4, Appli
33 87.5 12.3 2101 3 US-09-452-294-1 Sequence 1, Appli
34 87.5 12.3 2101 5 PCT-US93-08160-4 Sequence 4, Appli
35 87 12.2 1093 5 PCT-US93-03077-1 Sequence 1, Appli
36 87 12.2 1388 2 US-08-685-576-4 Sequence 4, Appli
37 86.5 12.2 710 4 US-09-107-532A-5067 Sequence 5067, Ap
38 86.5 12.2 2285 4 US-08-308-375-2 Sequence 2, Appli
39 86 12.1 1388 2 US-08-685-576-1 Sequence 1, Appli
40 85.5 12.0 410 1 US-08-471-033-40 Sequence 40, Appl
41 85.5 12.0 410 1 US-08-471-033-43 Sequence 43, Appl
42 85.5 12.0 410 2 US-08-471-044-40 Sequence 40, Appl
43 85.5 12.0 410 2 US-08-471-044-43 Sequence 43, Appl
44 85.5 12.0 410 2 US-08-463-483A-40 Sequence 40, Appl
45 85.5 12.0 410 2 US-08-463-483A-43 Sequence 43, Appl

ALIGNMENTS

RESULT 1
US-09-252-991A-19000
; Sequence 19000, Application US/09252991A
; Patent No. 6851795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19000
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19000

Query Match 16.8%; Score 119.5; DB 4; Length 165;
Best Local Similarity 25.9%; Pred. No. 0.00013;
Matches 35; Conservative 32; Mismatches 49; Indels 19; Gaps 5;
QY 1 MNISVNPYLMAVVFVVFVLLWMNVYRPLAFMNRQAEIKDSLAKIKTDNAQSVET 60
Db 8 VGVNINATLIGQSVAFVFLVFCMK-FWPPVIALQERQKXIADGL-DAANRAARDEL 65
QY 61 GHQIEALLKEAAEKRETIIEAIOKATESYDAVTKQENELNQEFFAFQOLQNEKQALK 120
Db 66 AHE-----KAGQQLREKAQAAE-----IVEQAKKANQIVDEARDQARTGERL 111
QY 121 EQLQACMPVFEDLN 135
Db 112 AQAQAEI---EQELN 123

RESULT 2
US-09-914-259-11
; Sequence 11, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259

; CURRENT FILING DATE: 2000-11-21
 ; NUMBER OF SEQ ID NOS: 180
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 11
 ; LENGTH: 3878
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-914-259-11

Query Match 15.2%; Score 108; DB 4; Length 3878;
 Best Local Similarity 31.0%; Pred. No. 0.1;
 Matches 31; Conservative 26; Mismatches 33; Indels 10; Gaps 4;
 QY 42 EIKDLSAKITD--NAQSVIEIGHQIEALLKEAAEK---RREIIARAIQKATESYDAVIKQ 96
 Db 716 LQQSLVNSSEMTQLINELQKEIILLQBEKEKGTLEQV--GELQLKTELLEKQWKE 773
 QY 97 KENELNQEF---EAFKQQLQNEKQALKQEQVMPFEDE 133
 Db 774 KENDLQEKFAQLEAENSILKDEKTKTLEMLKIHTPVSQEE 813

RESULT 3

US-09-134-001C-4229
 ; Sequence 4229, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 4229
 ; LENGTH: 178
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-4229

Query Match 15.1%; Score 107.5; DB 4; Length 178;
 Best Local Similarity 21.9%; Pred. No. 0.0023;
 Matches 30; Conservative 38; Mismatches 56; Indels 13; Gaps 3;
 QY 12 VVFWVFLLMANVWVYRPLLAFLMNRQAEI-----KDSLAKITDNAQSV-EI 60
 Db 28 VTLVTVILILKLFANGFLKVMVKREDINDKIDDAEQAKINAQKLEENRKLKET 87
 QY 61 GHQIEALLKEAAEKREIIAEAIQKATESYDAVIKQENELNOEFAKQLONEKQALK 120
 Db 88 QDEVKILDDAKIQARKQHEEIIHEANEKANGMITAQSEINSQKRAISINNQSVELS 147
 QY 121 EQLOQMPVFEDELNR 137
 Db 148 VLIASK--VLRKEISEQ 162

RESULT 4

US-09-328-352-4908
 ; Sequence 4908, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 4908
 ; LENGTH: 159
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-4908

Query Match 14.6%; Score 103.5; DB 4; Length 159;
 Best Local Similarity 25.7%; Pred. No. 0.005; 54; Indels 27; Gaps 6;
 Matches 38; Conservative 29; Mismatches 54; Indels 27; Gaps 6;
 QY 1 MNISVNYLMAVVFVYVLLWAMVWVYRPLLAFLMNRQAEIKDSL---AKIKTDNAQS 57
 Db 4 MNINUTLIGQAIATAFAFFVAFCKM---FVWPLINAISERQKRIADGLNAAEKAKADLADA 60
 QY 58 -VEIGHQIEA-----LLKEAAEKREIIAE-----IQKATESYDAVIKQKE 98
 Db 61 QACVQKQELDAAKAAQAQALIEQANRRRAQOLIEBATQAAEGERIRQQAKEAVDOEINSAR 120
 QY 99 NELNQEFEPFAKQLONEKQALKQEQVMPFEDE 136
 Db 121 EELRQOVAALA--VTGAEKILNQVDAK 146

RESULT 5

US-09-134-001C-2916
 ; Sequence 2916, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 2916
 ; LENGTH: 563
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-2916

Query Match 13.8%; Score 98; DB 4; Length 563;
 Best Local Similarity 26.8%; Pred. No. 0.088;
 Matches 26; Conservative 30; Mismatches 33; Indels 8; Gaps 4;
 QY 47 LAKIKTDNAQSVIEIGHQIEALLKEAAEKRE---IIAEAIQKATESYDAVIKQENELNQ 103
 Db 388 IEELKQKHSNOTEIKYDIDSLEKQKAKLKQOERLELEFLDDOMDS--GMLKAKQSEMNO 445
 QY 104 EPEAFKQLONEKQALKQEQVMPFEDELNKEVAM 140
 Db 446 QLEVDQIQIEAKQA--NQSQDEIPNP-DRLKGRLL 479

RESULT 6

US-09-914-259-66
 ; Sequence 66, Application US/09914259
 ; Patent No. 6495336
 ; GENERAL INFORMATION:
 ; APPLICANT: Makowski, Lee
 ; APPLICANT: Hyman, Paul
 ; APPLICANT: Williams, Mark
 ; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
 ; FILE REFERENCE: 8471-010-999
 ; CURRENT APPLICATION NUMBER: US/09/914,259
 ; CURRENT FILING DATE: 2000-11-21
 ; NUMBER OF SEQ ID NOS: 180
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 66

LENGTH: 961
TYPE: PRT
ORGANISM: Bos taurus
US-09-914-259-66

Query Match 13.2%; Score 94; DB 4; Length 961;
Best Local Similarity 30.3%; Pred. No. 0.44; 30; Indels 32; Gaps 8;
Matches 37; Conservative 23; Mismatches 30; Indels 32; Gaps 8;
QY 37 DNRC-ABIKSLAKIKTD-NAQSVIEIGHQIEALKEAAKRRREIIAIAIAQK-----85
DB DSEQIAELKQELATLKSQNSQSVI-TKLQTEKQELLQK-----TEAFKSAFVPGESE 835
QY 86 ---ATESVD-----AVIKQENELNQEFPAFQKQNEKQALKEQLQAO---MPVFEDEL 134
DB TVIATKTVDVEGRSALLQETKELKEIKA-----LSERTAIKEQLDSSNSTIALLQNEK 891
QY 135 NK 136
DB 892 NK 893

RESULT 7
US-09-104-324B-4
Sequence 4, Application US/09104324B
Patent No. 6232460
GENERAL INFORMATION:
APPLICANT: T rec1, Ozlem, Sahin, Ugur; Pfreundschuh, Michael
TITLE OF INVENTION: Methods For Diagnosis And Treating Cancers,
TITLE OF INVENTION: And Methods For Identifying Pathogenic Markers In A Sample Of
TITLE OF INVENTION: And Methods For Identifying Pathogenic Markers In A Sample Of
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Fulbright & Jaworski LLP
CITY: New York City
STATE: New York
ZIP: 10103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/104,324B
FILING DATE: 25-June-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/892,702
FILING DATE: 15-July-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 6232460man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5491
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 318-3000
TELEFAX: (212) 752-5958
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 976 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-09-104-324B-4

Query Match 12.9%; Score 92; DB 3; Length 976;
Best Local Similarity 25.2%; Pred. No. 0.7; 34; Indels 16; Gaps 4;
Matches 28; Conservative 33; Mismatches 34; Indels 16; Gaps 4;
QY 36 MDNRQABTKDS-----LAKIKTDNAQSVIEIGHQIEALKEAAKRRREIIAIAIAQKATE 86
DB 541 LKNOQEDINNNKQOERMLKQIENLQETOTLRNELEYVREELKQKDEVKCK-LDKSEE 599
QY 89 SYDAVIKQENELNQEFPAFQKQNEKQALKEQLQAO---QMPVFEDELNK 136

DB 600 NCNNLRQVENK-----NKYIEELQENKALKKKGTABSKQLNVIKYNK 645
RESULT 8
US-09-328-352-5515
Sequence 5515, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 5515
LENGTH: 208
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-5515

Query Match 12.8%; Score 91; DB 4; Length 208;
Best Local Similarity 23.3%; Pred. No. 0.12; 47; Indels 8; Gaps 1;
Matches 24; Conservative 24; Mismatches 47; Indels 8; Gaps 1;
QY 41 AEIKDSLAK-----IKTDNAQSVIEIGHQIEALKEAAKRRREIIAIAIAQKATESYDA 92
DB 39 AEAKEAVAKTIRAKKAVGTAKQVVEVAEKTEQVKEVAHEATSOVKAEIAETQOVKA 98
QY 93 VIKQENELNQEFPAFQKQNEKQALKEQLQAOQMPVFEDELN 135
DB 99 VINTEEKIEAETQELNQNIHQFAQIKQDILQRLDVIKRAQFN 141

RESULT 9
US-09-336-447A-13
Sequence 13, Application US/09336447A
Patent No. 6310190
GENERAL INFORMATION:
APPLICANT: HANSEN, ERIC J.
APPLICANT: AEBI, CHRISTOPH
APPLICANT: COPE, LESLIE D.
APPLICANT: MACIVER, ISOBEL
APPLICANT: FISKE, MICHAEL J.
APPLICANT: FREDENBURG, ROSS A.
TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
FILE REFERENCE: AMCY:024
CURRENT APPLICATION NUMBER: US/09/336,447A
CURRENT FILING DATE: 1999-06-21
NUMBER OF SEQ ID NOS: 98
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 13
LENGTH: 873
TYPE: PRT
ORGANISM: Moraxella catarrhalis
US-09-336-447A-13

Query Match 12.8%; Score 91; DB 4; Length 873;
Best Local Similarity 29.4%; Pred. No. 0.77; 44; Indels 26; Gaps 7;
Matches 37; Conservative 19; Mismatches 44; Indels 26; Gaps 7;
QY 42 EIKDSLAKIKTDNAQSV-----EIGHQ-----IEALLKEAAKRRREII---AEA---IQ 84
DB 610 ENKDGIAKNQADIANNIKVIYELAAQQDDQHSDDIKTLAKVSAANTDRIAKNKAADASPE 669
QY 85 KATESYDAVTKQKE-----NELNQEFPAFQKQ-NEKQALKEQ---LQAOQMPVFEDELN 135
DB 670 TLTKNQNTLIEQGEALVEQNKAINQELGPAHADVDQKQILQADITAKTAEIQNIN 729
QY 136 KRVANG 141
DB 730 RTVANG 735

```

RESULT 10
US-08-736-770-1
; Sequence 1, Application US/08736770
; Patent No. 5871965
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Au-Young, Janice
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NOVEL HUMAN GUANYLATE BINDING PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/736,770
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0145 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 608 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: Consensus
;
US-08-736-770-1

```

```

Query Match 12.7%; Score 90.5; DB 2; Length 608;
Best Local Similarity 23.8%; Pred. No. 0.54;
Matches 29; Conservative 34; Mismatches 38; Indels 21; Gaps 6;

QY 32 LIAFMDNRQAEIKDSIAKTKTNAQS-VEIGHQIEALLKEAAEKRRRIIEAIOKATE-- 88
DB 460 LKYLESKK-DVADAL--LQDQSLSEKEKAIEVERIKAESAEAKKWL-EEIQKNEEM 515
QY 89 -----SYDAVIKQ-----KENLNQEFFAFAKQLQNEKQALKQEQVFEDEL 134
DB 516 MDQKESYQEHVKQLTKMERDRQAQMEEQKTLTSLIQEQARKERCQESTQLQNEI 575
QY 135 NK 136
DB 576 QK 577

RESULT 11
US-08-630-822A-62
; Sequence 62, Application US/08630822A
; Patent No. 5840695
; GENERAL INFORMATION:
; APPLICANT: FRANK, GLENN R.
; APPLICANT: HUNTER, SHIRLEY WU

```

```

; APPLICANT: WALLENFELS, LYNDA
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,822A
; FILING DATE: 11-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CONNELL, GARY J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-17-C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 900 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-630-822A-62

Query Match 12.7%; Score 90.5; DB 2; Length 900;
Best Local Similarity 24.1%; Pred. No. 0.9;
Matches 26; Conservative 22; Mismatches 33; Indels 27; Gaps 3;

QY 33 LAFMDNRQAEIKDSIAKTKT-----DNAQSVEIGHQIEALLKEAAEKRRRIIEAIOKATE 88
DB 753 VAHURDAKRVNEEELHKLTKARSVDNAQMKELQEQVE-----AQQ 792
QY 89 SYDAVIKQKENLNQEFFAFAKQLQ---NEKQALKQEQVFEDEL 133
DB 793 VFSTLYKTHSNELKEELKSRHIQEMEERESILVHQLQIALARADSE 840

RESULT 12
US-09-005-069-62
; Sequence 62, Application US/09005069
; Patent No. 5932470
; GENERAL INFORMATION:
; APPLICANT: FRANK, GLENN R.
; APPLICANT: HUNTER, SHIRLEY WU
; APPLICANT: WALLENFELS, LYNDA
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/005,069

```

```
/
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION NUMBER: 08/630,822
/ APPLICATION NUMBER: 08/630,822
/ FILING DATE: 11-APR-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: CONNELL, GARY J.
/ REGISTRATION NUMBER: 32,020
/ REFERENCE/DOCKET NUMBER: 2618-17-C3
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (303) 863-9700
/ TELEFAX: (303) 863-0223
/ INFORMATION FOR SEQ ID NO: 62:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 900 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-09-005-069-62

Query Match 12.7%; Score 90.5; DB 2; Length 900;
Best Local Similarity 24.1%; Pred. No. 0.9;
Matches 26; Conservative 22; Mismatches 33; Indels 27; Gaps 3;

QY 33 LAPMNRQAEIKDSLAKIKT----DNAQSVFIGHQIEALKEAAEKREIIAEAIQKATE 88
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 753 VAHLRDAKRNVEEELHKLKTARSDNAQMKELQEQVE-----AEQ 792

QY 89 SYDAVIKQKENELNQEFAPAKQLQ---NEKQALKEQLQACMPVFEDF 133
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 793 VFSTLYKTHSNELKEEKSRIHQEMEEERESLVHQLQIALARADSE 840

RESULT 13
US-09-171-156A-21
; Sequence 21, Application US/09171156A
; Patent No. 6368845
; GENERAL INFORMATION:
; APPLICANT: Hunter, Shirley Wu
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS AND APPARATUS TO COLLECT SUCH PROTEINS
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SHERIDAN ROSS P.C.
; STREET: 1560 BROADWAY, SUITE 1200
; CITY: DENVER
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/171,156A
; FILING DATE: 04-Mar-1999
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-17-C4-PUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/863-9700
; TELEFAX: 303/863-0223
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 900 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
/
```

```
/
/ SEQUENCE DESCRIPTION: SEQ ID NO: 21:
/ US-09-171-156A-21

Query Match 12.7%; Score 90.5; DB 4; Length 900;
Best Local Similarity 24.1%; Pred. No. 0.9;
Matches 26; Conservative 22; Mismatches 33; Indels 27; Gaps 3;

QY 33 LAPMNRQAEIKDSLAKIKT----DNAQSVFIGHQIEALKEAAEKREIIAEAIQKATE 88
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 753 VAHLRDAKRNVEEELHKLKTARSDNAQMKELQEQVE-----AEQ 792

QY 89 SYDAVIKQKENELNQEFAPAKQLQ---NEKQALKEQLQACMPVFEDF 133
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 793 VFSTLYKTHSNELKEEKSRIHQEMEEERESLVHQLQIALARADSE 840

RESULT 14
US-09-004-730A-21
; Sequence 21, Application US/09004730A
; Patent No. 6485968
; GENERAL INFORMATION:
; APPLICANT: Weber, Eric
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Sim, Gek-Kee
; APPLICANT: Frank, Glenn
; APPLICANT: Wallenfels, Lynda
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS AND APPARATUS TO COLLECT SUCH PROTEINS
; FILE REFERENCE: 2618-17-C5-PUS-1
; CURRENT APPLICATION NUMBER: US/09/004,730A
; CURRENT FILING DATE: 1998-01-08
; PRIOR APPLICATION NUMBER: PCT/97US/18669
; PRIOR FILING DATE: 1997-10-15
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 21
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Ctenocephalides felis
/ US-09-004-730A-21

Query Match 12.7%; Score 90.5; DB 4; Length 900;
Best Local Similarity 24.1%; Pred. No. 0.9;
Matches 26; Conservative 22; Mismatches 33; Indels 27; Gaps 3;

QY 33 LAPMNRQAEIKDSLAKIKT----DNAQSVFIGHQIEALKEAAEKREIIAEAIQKATE 88
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 753 VAHLRDAKRNVEEELHKLKTARSDNAQMKELQEQVE-----AEQ 792

QY 89 SYDAVIKQKENELNQEFAPAKQLQ---NEKQALKEQLQACMPVFEDF 133
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 793 VFSTLYKTHSNELKEEKSRIHQEMEEERESLVHQLQIALARADSE 840

RESULT 15
US-08-981-799A-21
; Sequence 21, Application US/08981799A
; Patent No. 6576238
; GENERAL INFORMATION:
; APPLICANT: Weber, Eric
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Sim, Gek-Kee
; APPLICANT: Frank, Glenn
; APPLICANT: Wallenfels, Lynda
; TITLE OF INVENTION: "NOVEL ECTOPARASITE SALIVA PROTEINS AND APPARATUS TO COLLECT SUCH PROTEINS"
; FILE REFERENCE: 2618-17-C5-PUS
; CURRENT APPLICATION NUMBER: US/08/981,799A
; CURRENT FILING DATE: 1998-08-27
; PRIOR APPLICATION NUMBER: PCT/97/18669
; PRIOR FILING DATE: 1997-10-15
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 21
/
```

```

; LENGTH: 900
; TYPE: PRT
; ORGANISM: Ctenocephalides felis
US-08-981-799A-21

Query Match      12.7%; Score 90.5; DB 4; Length 900;
Best Local Similarity 24.1%; Pred. No. 0.9;
Matches 26; Conservative 22; Mismatches 33; Indels 27; Gaps 3;

QY 33 LAFMDNQAEIKDSLAKIT-----DNAQSVETGCHQIEALLKEAAEKRRRIIAEAIQKATE 88
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 753 VAHLRDAKRNVEELHKLKTARSVDNAQMKELQEQVE-----AEQ 792

QY 89 SYDAVIKQKENELNQEFAPAKOLO--NEKQALKEOLOAQMEVFEDE 133
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 793 VFSTLYKTHSNELKEELEKSRHIQEMEERESLVHQLQIALARADSE 840

```

Search completed: December 10, 2003, 18:38:39
Job time : 10.0471 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:33:52 ; Search time 17.7173 Seconds
(without alignments)
1511.612 Million cell updates/sec

Title: US-10-080-113-4
Perfect score: 711
Sequence: 1 MNISVNPYLMAVVFVLL.....AQMPVFEELNKRVMGLGS 144

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 684280 segs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	711	100.0	144	10	US-09-881-752A-306
2	711	100.0	144	12	US-10-080-113-4
3	125	17.6	156	9	US-09-815-242-14068
4	123	17.3	156	9	US-09-741-669-340
5	123	17.3	156	9	US-09-815-242-10402
6	120.5	16.9	156	9	US-09-815-242-12114
7	108	15.2	160	9	US-09-864-761-47959
8	108	15.2	820	12	US-10-029-386-32324
9	108	15.2	3878	12	US-10-080-608A-11
10	108	15.2	3899	15	US-10-171-311-4
11	108	15.2	3907	15	US-10-171-311-2
12	108	15.2	3911	12	US-10-370-685-100
13	108	15.2	3917	15	US-10-171-311-8
14	108	15.2	3925	15	US-10-171-311-6
15	102	14.3	1001	15	US-10-128-714-3240

Sequence 191, App	945	13.9	98.5	16	US-09-745-763-191
Sequence 1540, App	1000	13.9	98.5	17	US-10-012-697-1540
Sequence 219, App	1711	13.9	98.5	18	US-09-771-161A-219
Sequence 220, App	1711	13.9	98.5	19	US-09-771-161A-220
Sequence 4838, App	188	13.7	97.5	20	US-09-738-626-4838
Sequence 2, Appli	1719	13.6	97	21	US-10-288-798-2
Sequence 5440, App	168	13.4	95.5	22	US-09-815-242-5440
Sequence 12386, A	173	13.4	95.5	23	US-09-815-242-12386
Sequence 12822, A	173	13.4	95.5	24	US-09-815-242-12822
Sequence 13135, A	173	13.4	95.5	25	US-09-815-242-13135
Sequence 7646, App	1881	13.4	95	26	US-10-032-585-7646
Sequence 66, Appl	961	13.2	94	27	US-10-080-608A-66
Sequence 155, App	961	13.2	94	28	US-10-370-685-155
Sequence 104, App	990	13.2	94	29	US-10-074-511-104
Sequence 32, Appl	312	13.2	93.5	30	US-09-976-782-32
Sequence 351, App	216	13.0	92.5	31	US-10-043-487-351
Sequence 596, App	976	12.9	92	32	US-10-117-937-596
Sequence 164, App	1938	12.9	91.5	33	US-10-171-311-164
Sequence 2, Appli	1945	12.9	91.5	34	US-09-927-597-2
Sequence 103, App	1972	12.9	91.5	35	US-10-341-434-103
Sequence 162, App	1972	12.9	91.5	36	US-10-171-311-162
Sequence 4, Appli	1979	12.9	91.5	37	US-09-927-597-4
Sequence 13, Appl	873	12.8	91	38	US-09-952-267-13
Sequence 32003, A	253	12.7	90.5	39	US-10-029-386-32003
Sequence 21, Appl	300	12.7	90.5	40	US-10-071-751-21
Sequence 1617, A	1045	12.7	90	41	US-09-815-242-10617
Sequence 73, Appl	1805	12.7	90	42	US-09-820-843A-73
Sequence 305, App	689	12.5	89	43	US-10-108-605-305
Sequence 6, Appli	471	12.4	88.5	44	US-10-038-686-6
Sequence 625, App	645	12.4	88.5	45	US-09-764-868-625

ALIGNMENTS

RESULT 1
US-09-881-752A-306
; Sequence 306, Application US/09881752A
; Patent No. US20030115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881.752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 306
; LENGTH: 144
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-881-752A-306

Query Match	100.0%	Score 711:	DB 10:	Length 144;
Best Local Similarity	100.0%	Pred. No. 3.7e-57;		
Matches 144;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MNISVNPYLMAVVFVLLWMNVVYRPLLAFLMNDNRQAEIKDSIAKIKTDAQSVET	60	
Db	1	MNISVNPYLMAVVFVLLWMNVVYRPLLAFLMNDNRQAEIKDSIAKIKTDAQSVET	60	
QY	61	GHOIALLKEAAKREIRIIAIAIKATSYDAVIKQENELNQEFQAFQLOKQKALK	120	
Db	61	GHOIALLKEAAKREIRIIAIAIKATSYDAVIKQENELNQEFQAFQLOKQKALK	120	

Qy 121 EQLQAMPVFEDELNKRKRVAMGLGS 144
Db 121 EQLQAMPVFEDELNKRKRVAMGLGS 144

RESULT 2
US-10-080-113-4
; Sequence 4, Application US/10080113
; Publication No. US20030166027A1
; GENERAL INFORMATION:
; APPLICANT: SACHS, GEORGE
; APPLICANT: VOLAND, BETRA
; TITLE OF INVENTION: COMPOSITIONS, TEST KITS AND METHODS FOR DETECTING
; TITLE OF INVENTION: HELICOBACTER PYLORI
; FILE REFERENCE: 626 06 PA
; CURRENT APPLICATION NUMBER: US/10/080,113
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 144
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-10-080-113-4

Query Match 100.0%; Score 711; DB 12; Length 144;
Best Local Similarity 100.0%; Pred. No. 3.7e-57;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MNISVNPYLMAVVFVLLWMNVVYRPLAFMDNRQAEIKDSLAKIKTDNAQSVEI 60
Db 1 MNISVNPYLMAVVFVLLWMNVVYRPLAFMDNRQAEIKDSLAKIKTDNAQSVEI 60
Qy 61 GHQIEALLKEAEKREIRIIAEAIQKATESYDAVIKQENELNQEFAFAKOLNEKQALK 120
Db 61 GHQIEALLKEAEKREIRIIAEAIQKATESYDAVIKQENELNQEFAFAKOLNEKQALK 120
Qy 121 EQLQAMPVFEDELNKRKRVAMGLGS 144
Db 121 EQLQAMPVFEDELNKRKRVAMGLGS 144

RESULT 3
US-09-815-242-14068
; Sequence 14068, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/131,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14068
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(156)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-14068

Query Match 17.6%; Score 125; DB 9; Length 156;
Best Local Similarity 28.5%; Pred. No. 0.00066;
Matches 41; Conservative 30; Mismatches 51; Indels 22; Gaps 5;
Qy 1 MNISVNPYLMAVVFVLLWMNVVYRPLAFMDNRQAEIKDSLAKIKTDNAQSVEI 60
Db 1 MNLNATIXGQATAFILFV---WFCMKYVYVPPMLAAIEKROKEIADGLA-----SAER 49
Qy 61 GHQIEALLKEAEKREIRIIAEAIQKATESYDAVIKQENELNQEFAFAKOLNEK 116
Db 50 AHKDLDLAKASATDQKKAKAEAOVIEQANKRRAQILDEAKTEAEQERTKIVACAQAEI 109
Qy 117 QALKEQLQAMPVFEDELNKRKRVAM 140
Db 110 EAERKEAR-----EELRKQVAI 126

RESULT 4
US-09-741-669-340
; Sequence 340, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; TITLE OF INVENTION: Proliferation of E. coli
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 340
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-669-340

Query Match 17.3%; Score 123; DB 9; Length 156;
Best Local Similarity 27.1%; Pred. No. 0.001;
Matches 39; Conservative 27; Mismatches 56; Indels 22; Gaps 4;
Qy 1 MNISVNPYLMAVVFVLLWMNVVYRPLAFMDNRQAEIKDSLAKIKTDNAQSVEI 60
Db 1 MNLNATIXGQATAFILFVFCMK---YVPPMLAAIEKROKEIADGLA-----SAER 49
Qy 61 GHQIEALLKEAEKREIRIIAEAIQKATESYDAVIKQENELNQEFAFAKOLNEKQALK 120
Db 50 AHKDLDLAKASATDQKKAKAEAOVIEQANKRRAQILDEAKTEAEQERTKIVACAQAEI 102
Qy 121 EQLQAMPV-----FEDELNKRKRVAM 140
Db 103 AQAQAEIEAEKREIRIAREELRKQVAI 126

RESULT 5
US-09-815-242-10402
; Sequence 10402, Application US/09815242

```
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10402
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10402

Query Match          17.3%; Score 123; DB 9; Length 156;
Best Local Similarity 27.1%; Pred. No. 0.001;
Matches 39; Conservative 27; Mismatches 56; Indels 22; Gaps 4;

QY 1 MNISVNPYLMVAVFVFWVLLWMNVVYRLLAFMNRQAEIKDSLAKIKTNAQSVEI 60
DB 1 MNLNATLIGQIAFVFLVFCMK---YVWPPPLMAAEIKRQKEIADGLA-----SAER 49
QY 61 GHOIEALLKEAAEKREIRIIAIAIQKATESYDAVIKQKNEINQEFFAFAKQKQALK 120
DB 50 AEKLDLAKASA-----TQLKAKAEACQVILIEQANKRSQILDKAEAEQCRTKIV 102
QY 121 EQLQKQMPV-----FEDELNKRVM 140
DB 103 AQAQAEIEAEKRAFEELRKQVAI 126

RESULT 6
US-09-815-242-12114
; Sequence 12114, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
```

```
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12114
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12114

Query Match          16.9%; Score 120.5; DB 9; Length 156;
Best Local Similarity 28.1%; Pred. No. 0.0017;
Matches 38; Conservative 27; Mismatches 49; Indels 21; Gaps 5;

QY 1 MNISVNPYLMVAVFVFWVLLWMNVVYRLLAFMNRQAEIKDSLAKIKTNAQSVEI 60
DB 1 MNLNATLIGQIAFVFLVFCMK---FWPFPVIAAQERKQKIADGL-DANRAARDEL 56
QY 61 GHOIEALLKEAAEKREIRIIAIAIQKATESYDAVIKQKNEINQEFFAFAKQKQALK 120
DB 57 AHE-----KAGQQLREAKQAAB-----IVEQAKKRAQIVDEARDQARTGERLK 102
QY 121 EQLQKQMPVFEDELN 135
DB 103 AQAQAEI---EQELN 114

RESULT 7
US-09-864-761-47959
; Sequence 47959, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
```

; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00660
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-29
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 47959
 ; LENGTH: 660
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AJ010770.1
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
 ; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
 ; OTHER INFORMATION: SWISSPROT HIT: Q99323, EVALUE 3.00e-17
 ; OTHER INFORMATION: EST_HUMAN HIT: AUL32932.1, EVALUE 1.00e-105
 US-09-864-761-47959

Query Match 15.2%; Score 108; DB 9; Length 660;
 Best Local Similarity 31.0%; Pred. No. 0.14;
 Matches 31; Conservative 26; Mismatches 33; Indels 10; Gaps 4;
 QY 42 EIKDSLAKITD--NAQSVGHQIEALLKEAAEK---RRRIIAEAIQKATESYDAVIKQ 96
 DB 394 DLQSLVNSKSEMTLQINELQKEIEILRQKEKGTLEQEV--QELQKTELLEKQMK 451
 QY 97 KENELNQEFF--EAFKQLQNEKQALKEQLQAOQMPVFEDE 133
 DB 452 KENDLQKFAQLAEASILKDEKKTLEDMLKIHTFPVSQEE 491

RESULT 8
 US-10-029-386-32324
 ; Sequence 32324, Application US/10029386
 ; Publication No. US20030194704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Hanzel, David K.
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
 ; FILE REFERENCE: AEMICA-X-2
 ; CURRENT APPLICATION NUMBER: US/10/029,386
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 34288
 ; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 32324
 ; LENGTH: 820
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AC000066.1
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.87
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
 ; OTHER INFORMATION: SWISSPROT HIT: Q99996, EVALUE 0.00e-00
 US-10-029-386-32324

Query Match 15.2%; Score 108; DB 12; Length 820;

Best Local Similarity 31.0%; Pred. No. 0.18;
 Matches 31; Conservative 26; Mismatches 33; Indels 10; Gaps 4;
 QY 42 EIKDSLAKITD--NAQSVGHQIEALLKEAAEK---RRRIIAEAIQKATESYDAVIKQ 96
 DB 410 DLQSLVNSKSEMTLQINELQKEIEILRQKEKGTLEQEV--QELQKTELLEKQMK 467
 QY 97 KENELNQEFF--EAFKQLQNEKQALKEQLQAOQMPVFEDE 133
 DB 468 KENDLQKFAQLAEASILKDEKKTLEDMLKIHTFPVSQEE 507

RESULT 9
 US-10-080-608A-11
 ; Sequence 11, Application US/10080608A
 ; Publication No. US20030198956A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Makowski, Lee
 ; APPLICANT: Hyman, Paul
 ; APPLICANT: Williams, Mark
 ; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
 ; FILE REFERENCE: 8471-010-999
 ; CURRENT APPLICATION NUMBER: US/10/080,608A
 ; CURRENT FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 180
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 11
 ; LENGTH: 3878
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-080-608A-11

Query Match 15.2%; Score 108; DB 12; Length 3878;
 Best Local Similarity 31.0%; Pred. No. 1.3;
 Matches 31; Conservative 26; Mismatches 33; Indels 10; Gaps 4;
 QY 42 EIKDSLAKITD--NAQSVGHQIEALLKEAAEK---RRRIIAEAIQKATESYDAVIKQ 96
 DB 716 DLQSLVNSKSEMTLQINELQKEIEILRQKEKGTLEQEV--QELQKTELLEKQMK 773
 QY 97 KENELNQEFF--EAFKQLQNEKQALKEQLQAOQMPVFEDE 133
 DB 774 KENDLQKFAQLAEASILKDEKKTLEDMLKIHTFPVSQEE 813

RESULT 10
 US-10-171-311-4
 ; Sequence 4, Application US/10171311
 ; Publication No. US20030087270A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Schlegel, Robert
 ; APPLICANT: Chen, Fan
 ; APPLICANT: Zhao, Xumei
 ; APPLICANT: Monahan, John
 ; APPLICANT: Kamatkar, Shubhangi
 ; APPLICANT: Glatt, Karen
 ; APPLICANT: Ganavarapu, Manjula
 ; APPLICANT: Hoersch, Sebastian
 ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
 ; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
 ; TITLE OF INVENTION: OF CERVICAL CANCER
 ; FILE REFERENCE: MRI-035
 ; CURRENT APPLICATION NUMBER: US/10/171,311
 ; CURRENT FILING DATE: 2002-06-12
 ; PRIOR APPLICATION NUMBER: US 60/298,159
 ; PRIOR FILING DATE: 2001-06-13
 ; PRIOR APPLICATION NUMBER: US 60/298,155
 ; PRIOR FILING DATE: 2001-06-13
 ; PRIOR APPLICATION NUMBER: US 60/335,936
 ; PRIOR FILING DATE: 2001-11-14
 ; NUMBER OF SEQ ID NOS: 238
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4

```
; LENGTH: 3899
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-4

Query Match
Best Local Similarity 15.2%; Score 108; DB 15; Length 3899;
Matches 31; Conservative 26; Mismatches 33; Indels 10; Gaps 4;

QY 42 EIKDSLAKIKTD--NAQSVFCHQIEALLKEAAEK---RREIIAEAIQKATESYDAVIKQ 96
Db 704 DIQQSLVNSKSEMTLQINELQKEIEILRQEEKGTLEQEV--QELQLKTELLEKQMK 761

QY 97 KENELNQEF---EAFKQLONEKQALKEQLQAOQMPVFEDE 133
Db 762 KENDLQEKFAQLEAENSILKDEKKTLEDMLKIHTPVSQEE 801

RESULT 11
US-10-171-311-2
; Sequence 2, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhang
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 3907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-2

Query Match
Best Local Similarity 15.2%; Score 108; DB 15; Length 3907;
Matches 31; Conservative 26; Mismatches 33; Indels 10; Gaps 4;

QY 42 EIKDSLAKIKTD--NAQSVFCHQIEALLKEAAEK---RREIIAEAIQKATESYDAVIKQ 96
Db 704 DIQQSLVNSKSEMTLQINELQKEIEILRQEEKGTLEQEV--QELQLKTELLEKQMK 761

QY 97 KENELNQEF---EAFKQLONEKQALKEQLQAOQMPVFEDE 133
Db 762 KENDLQEKFAQLEAENSILKDEKKTLEDMLKIHTPVSQEE 801

RESULT 12
US-10-370-685-100
; Sequence 100, Application US/10370685
; Publication No. US20030215903A1
; GENERAL INFORMATION:
; APPLICANT: Hyman, Paul
; APPLICANT: Goldberg, Edward
; TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements
; FILE REFERENCE: NANF.P-004

; CURRENT APPLICATION NUMBER: US/10/370,685
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: 10/080,608
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 100
; LENGTH: 3911
; TYPE: PRT
; ORGANISM: human
US-10-370-685-100

Query Match
Best Local Similarity 15.2%; Score 108; DB 12; Length 3911;
Matches 31; Conservative 26; Mismatches 33; Indels 10; Gaps 4;

QY 42 EIKDSLAKIKTD--NAQSVFCHQIEALLKEAAEK---RREIIAEAIQKATESYDAVIKQ 96
Db 716 DIQQSLVNSKSEMTLQINELQKEIEILRQEEKGTLEQEV--QELQLKTELLEKQMK 773

QY 97 KENELNQEF---EAFKQLONEKQALKEQLQAOQMPVFEDE 133
Db 774 KENDLQEKFAQLEAENSILKDEKKTLEDMLKIHTPVSQEE 813

RESULT 13
US-10-171-311-8
; Sequence 8, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhang
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 3917
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-8

Query Match
Best Local Similarity 15.2%; Score 108; DB 15; Length 3917;
Matches 31; Conservative 26; Mismatches 33; Indels 10; Gaps 4;

QY 42 EIKDSLAKIKTD--NAQSVFCHQIEALLKEAAEK---RREIIAEAIQKATESYDAVIKQ 96
Db 704 DIQQSLVNSKSEMTLQINELQKEIEILRQEEKGTLEQEV--QELQLKTELLEKQMK 761

QY 97 KENELNQEF---EAFKQLONEKQALKEQLQAOQMPVFEDE 133
Db 762 KENDLQEKFAQLEAENSILKDEKKTLEDMLKIHTPVSQEE 801

RESULT 14
US-10-171-311-6
```

```

; Sequence 6, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Ganavarapu, Manjula
; APPLICANT: Hoersch, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 3925
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-6

Query Match      15.2%; Score 108; DB 15; Length 3925;
Best Local Similarity 31.0%; Pred. No. 1.3;
Matches 31; Conservative 26; Mismatches 33; Indels 10; Gaps 4;

QY 42 EIKDSLAKIKTD--NAQSVZIGHQIEALLKEAAEK---RREIARAIQXATESYDAVIKQ 96
DB 704 DLOQSLVNSSEMTLQINLQKEIILRQEKEXKTELEQEV--QELQKTELEKQMK 761
QY 97 KENELNQEF---EAFKQLQNEKQALKEQLQAQMPVFEDE 133
DB 762 KENDLQEKFAQLAENSILKDEKTKLEDMLKIHTPVSQEE 801

RESULT 15
US-10-128-714-3240
; Sequence 3240, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/295,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3240

```

```

; LENGTH: 1001
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3240

Query Match      14.3%; Score 102; DB 15; Length 1001;
Best Local Similarity 31.0%; Pred. No. 0.83;
Matches 27; Conservative 17; Mismatches 35; Indels 8; Gaps 2;

QY 37 DNRQAEIKDSLAKIKTDNAQSVZIGHQIEALLKEAAEKREIIRAEAIQKATESYDAVIKQ 96
DB 189 DTRQAEVDDVKAKLAESSENTIAKAGEAAEKLAELKQKTEI--EKQ-----DQVTKQ 240
QY 97 KENELNQEFQFAKQLQNEKQALKEQL 123
DB 241 KNDQOQKELQEKLEEVAKEDASEKKL 267

Search completed: December 10, 2003, 18:40:33
Job time : 18.7173 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:30:22 ; Search time 9.61257 Seconds
(without alignments)
1440.645 Million cell updates/sec

Title: US-10-080-113-4

Perfect score: 711

Sequence: 1 MNISVNPYLMAVFWVFVLL.....AQMVPFDELNKRVMGLGS 144

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	711	100.0	144	2 A64662	ATP synthase F0, s
2	685	96.3	144	2 A71856	ATP synthase b', -
3	169	23.8	141	2 R81426	H+-transporting tw
4	146	20.5	219	2 T05402	H+-transporting tw
5	143	20.1	159	2 S36962	H+-transporting tw
6	133	18.7	222	2 S34473	H+-transporting tw
7	132	18.6	1179	2 F71190	probable chromosom
8	127	17.9	164	2 R72231	ATP synthase F0, s
9	125	17.6	156	2 AB0954	ATP synthase chain
10	123	17.3	156	1 LWECEB	H+-transporting tw
11	123	17.3	156	2 F91213	membrane-bound ATP
12	123	17.3	156	2 G86059	ATP synthase subun
13	122.5	17.2	162	2 F84119	ATP synthase B cha
14	120.5	16.9	156	2 G82952	H+-transporting tw
15	120.5	16.9	163	2 S01399	H+-transporting tw
16	120	16.9	156	2 A10500	ATP synthase F0 su
17	119.5	16.8	144	1 C70437	ATP synthase, B ch
18	119.5	16.8	156	2 C82126	H+-transporting tw
19	119	16.7	188	1 S23223	Na+-transporting A
20	116.5	16.4	159	2 S26959	H+-transporting tw
21	116	16.3	158	2 S73166	H+-transporting tw
22	113.5	16.0	164	2 T06913	H+-transporting tw
23	112	15.8	164	2 H97702	H+-transporting tw
24	109.5	15.4	157	2 S39517	H+-transporting tw
25	108	15.2	1642	2 T08880	KMDA receptor-bind
26	107	15.0	156	1 S06078	H+-transporting tw
27	106	14.9	156	2 B82037	ATP synthase F0, B
28	106	14.9	161	2 D84930	H+-transporting tw
29	106	14.9	163	1 D31090	H+-transporting tw

ATP synthase chain
chromosome segrega
ATP synthase S cha
H+-transporting tw
H+-transporting tw
protein kinase (EC
microtubule bindin
H+-transporting tw
H+-transporting tw
H+-transporting tw
H+-transporting tw
myosin-like protei
H+-transporting tw
kinesin motor prot
hypothetical prote
hypothetical prote
H+-transporting tw

ALIGNMENTS

RESULT 1

A64662

ATP synthase F0, subunit b' - Helicobacter pylori (strain 26695)

C/Species: Helicobacter pylori

C/Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999

C/Accession: A64662

R/Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997

A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A/Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A/Reference number: A64520; MUID:97394467; PMID:9252185

A/Accession: A64662

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-144 <TOM>

A/Cross-references: GB:AE000619; GB:AE000511; NID:G9314276; PIDN:AAD08179.1; PID:G93142

Query Match

Best Local Similarity 100.0%; Score 711; DB 2; Length 144;

Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNISVNPYLMAVFWVFVLLWMNVVYRPLAFMDNRQAEIKDSLAKIKTDNAQSVEI 60
DB 1 MNISVNPYLMAVFWVFVLLWMNVVYRPLAFMDNRQAEIKDSLAKIKTDNAQSVEI 60

QY 61 GHQIEALLKEAAEKREIIAEAIQKATESYDAVIKQENELNQEFQAFKQLQNEKQALK 120
DB 61 GHQIEALLKEAAEKREIIAEAIQKATESYDAVIKQENELNQEFQAFKQLQNEKQALK 120

QY 121 EQLQAMPVFEDLNKRVMGLGS 144
DB 121 EQLQAMPVFEDLNKRVMGLGS 144

RESULT 2

A71856

ATP synthase b' - Helicobacter pylori (strain J99)

C/Species: Helicobacter pylori

A/Variety: strain J99

C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C/Accession: A71856

R/Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Werberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999

A/Title: Genomic sequence comparison of two unrelated isolates of the human gastric pat
A/Reference number: A71800; MUID:99120557; PMID:9923682

A/Accession: A71856

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-144 <ARN>

A;Cross-references: GB:AE001533; GB:AE001439; NID:g4155636; PIDN:AAD06630.1; PID:g415564
A;Experimental source: strain J99
C;Genetics:
A;Gene: atpX

Query Match 96.3%; Score 685; DB 2; Length 144;
Best Local Similarity 95.8%; Pred. No. 3.6e-40; Mismatches 4; Indels 0; Gaps 0;
Matches 138; Conservative 2;
Qy 1 MNISNPLYMAVVFVLLWAMNMYRPLAFMDNRQAEIKDSLAKIKTDNAQSVEI 60
Db 1 MNISNPLYMAVVFVLLWAMNMYRPLAFMDNRQAEIKDSLAKIKTDNTQSV 60
Qy 61 GHOIEALLKEAEKREIIAEAIQKATESYDAVIKQENELNORFEAFKOLQNEKQALK 120
Db 61 GHOIEALLKEAEKREIIAEAIQKATESYDAVIKQENELNORFEAFKOLQNEKQALK 120
Qy 121 EQIQAMQMPVFEDELNKRVMGLGS 144
Db 121 EQIQAMQMPVFEDELNKRVMGLGS 144

RESULT 3
F81426
H+-transporting two-sector ATPase (EC 3.6.3.14) F0 sector B' chain Cj0102 [imported] - C
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C;Accession: F81426
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A;Reference number: A81250; MUID:20150912; PMID:10698204
A;Accession: F81426
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-141 <PAR>
A;Cross-references: GB:AL139074; GB:AL111168; NID:g6967505; PIDN:CAB72586.1; PID:g696759
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: atpF; Cj0102
A;Keywords: hydrolase

Query Match 23.8%; Score 169; DB 2; Length 141;
Best Local Similarity 25.4%; Pred. No. 4.9e-05;
Matches 34; Conservative 37; Mismatches 63; Indels 0; Gaps 0;
Qy 5 VNPYLMAVVFVLLWAMNMYRPLAFMDNRQAEIKDSLAKIKTDNAQSVEIGHQI 64
Db 5 MHPSIMLATVAIFLAVIVILNSLYKPLKFMDEKNDISKDENKVKNSQEVLGWDEL 64
Qy 65 EALLKEAEKREIIAEAIQKATESYDAVIKQENELNORFEAFKOLQNEKQALKQ 124
Db 65 EATHINTREIIOIKOSATAAEAEQILRSKXELERKQASFYDLAVQKKELQEH 124
Qy 125 AQMPVFEDELNKR 138
Db 125 IHLSELKQALQNNI 138

RESULT 4
T05402
H+-transporting two-sector ATPase (EC 3.6.3.14) chain 9 - Arabidopsis thaliana
N;Alternate names: protein F10M6.100
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 03-Jun-2002
C;Accession: T05402
R;Sevan, M.; Weichselgartner, M.; Fartmann, B.; Granderath, K.; Dauner, D.; Herzl, A.; N
submitted to the Protein Sequence Database, February 1998
A;Reference number: Z15414
A;Accession: T05402
A;Molecule type: DNA
A;Residues: 1-219 <BEV>

A;Cross-references: EMBL:AL021811
A;Experimental source: cultivar Columbia; BAC clone F10M6
C;Genetics:
A;Map position: 4
A;Genome: nuclear
A;Intons: 165/3
A;Note: F10M6.100
C;Keywords: chloroplast; hydrolase; transmembrane protein

Query Match 20.5%; Score 146; DB 2; Length 219;
Best Local Similarity 25.6%; Pred. No. 0.0029;
Matches 34; Conservative 36; Mismatches 55; Indels 8; Gaps 2;
Qy 10 MAVVFVVFVLLWAMNMYRPLAFMDNRQAEIKDSLAKIKTDNAQSVEIGHQIEALLK 69
Db 88 LPIIVVEFTFLMFDLKVYSPFLGNFMDQDASIKELASVKDTSTEVKELDEQAAAVMR 147
Qy 70 EAAEKREIIAEAIQKATESYDAVIKQK-----ENELNORFEAFKOLQNEKQALKQ 125
Db 148 AA-----RABIAAALNKKKKTQVEVEEKLAEGKKVVEELKEALASLESQKEETIKALDS 203
Qy 126 QMPVFEDELNKR 138
Db 204 QIAALSEDIVKVV 216

RESULT 5
S36962
H+-transporting two-sector ATPase (EC 3.6.3.14) chain b' - Synechococcus sp. (PCC 6716)
C;Species: Synechococcus sp.
A;Variety: PCC 6716
C;Date: 22-Jan-1994 #sequence_revision 12-Apr-1996 #text_change 03-Jun-2002
C;Accession: S36962; S31886
R;van Walraven, H.S.; Lutter, R.; Walker, J.E.
Biochem. J. 294, 239-251, 1993
A;Title: Organization and sequences of genes for the subunits of ATP synthase in the th
A;Reference number: S36960; MUID:93371369; PMID:8363578
A;Accession: S36962
A;Molecule type: DNA
A;Residues: 1-159 <VAN>
A;Cross-references: EMBL:X70431; NID:g49213; PIDN:CAA49872.1; PID:g49216
A;Experimental source: PCC 6716
A;Superfamily: H+-transporting ATP synthase chain I
C;Keywords: ATP biosynthesis; hydrolase; membrane-associated complex

Query Match 20.1%; Score 143; DB 2; Length 159;
Best Local Similarity 27.7%; Pred. No. 0.0033;
Matches 36; Conservative 34; Mismatches 46; Indels 14; Gaps 2;
Qy 9 LMAVVFVVFVLLWAMNMYRPLAFMDNRQAEIKDSLAKIKTDNAQSVEIGHQIE--- 65
Db 31 LMAVQFILLTVIL---NALLYKPLGQALNDRDEYITNLQAKERLQQATELANQVEQEL 87
Qy 66 -----ALLKDAEKREIIAEAIQKATESYDAVIKQENELNORFEAFKOLQNEKQ 117
Db 88 AYTRREAQAIIEEAEAEQKIAETAAQAQALQALQALQALQALQALQALQALQALQ 147
Qy 118 ALKEQLQALQ 127
Db 148 SLSEQLLAKL 157

RESULT 6
S34473
H+-transporting two-sector ATPase (EC 3.6.3.14) chain 9 - spinach
C;Species: Spinacia oleracea (spinach)
C;Date: 08-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 03-Jun-2002
C;Accession: S34473
R;Herrmann, R.G.; Steppuhn, J.; Herrmann, G.S.; Nelson, N.
FEBS Lett. 326, 192-198, 1993
A;Title: The nuclear-encoded polypeptide Cfo-II from spinach is a real, ninth subunit of
A;Reference number: S34473; MUID:93314783; PMID:8325369
A;Accession: S34473

A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: H7231
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-164 <ARN>
A:Cross-references: GB:AE001805; GB:AE000512; NID:g4982180; PIDN:AAD36681.1; PID:g49821
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1614
C:Superfamily: H+-transporting ATP synthase chain I

Query Match 17.9%; Score 127; DB 2; Length 164;
Best Local Similarity 21.6%; Pred.No. 0.042;
Matches 32; Conservative 44; Mismatches 58; Indels 14; Gaps 4;

QY 3 ISVNPYLMAVVVFVILLWANNVVRPLLAFMDNRQAIEKDSLAKIKTKDNAQSVEIGH 62
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 4 LEIN-WTSAAMLMLFVLVMVFLNKFLYTPPIENAEKKRKVEDLSAEQLKSEAEKWS 62
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 63 QIBALUKAAEKREII-----AEAI-----QKATESYDAVIKQENELNQFEFAFKQ 111
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 63 EAERFLSEARQRADEIVESARKAEAVESAREKAKEAQNIIVESAQTQIEVYKKALEQ 122
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 112 LONEKQALKQELOQAQM--PVFEDELNKR 137
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 123 VQERRAELSUILATKLLOKFQDERARR 150
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

RESULT 9
ABO3954
ATP synthase chain B [imported] - Salmonella enterica subsp. enterica serovar Typhi (st
C:Species: Salmonella enterica subsp. enterica serovar Typhi
C>Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AB03954
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AB03954
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-156 <PAR>
A:Cross-references: GB:ALS13382; PIDN:CAD03126.1; PID:g16504762; GSPDB:GN00176
C:Genetics:
A:Gene: STY3909
C:Superfamily: H+-transporting ATP synthase chain I

Query Match 17.6%; Score 125; DB 2; Length 156;
Best Local Similarity 28.5%; Pred.No. 0.054;
Matches 41; Conservative 50; Mismatches 51; Indels 22; Gaps 5;

QY 1 MNISVNPLYMAVVVFVILLWANNVVRPLLAFMDNRQAIEKDSLAKIKTKDNAQSVEI 60
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1 MLNLNATILGQAIAFILFV--WFCMKYVPPLMAAIEKRQKEIADGLA-----SAER 49
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 61 GHQIEALLK-EAAEKRREITAEA---IQKATESYDAVIKQENELNQFEFAFAQLQNEK 116
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 50 AHKDLDLKASATDQLKKKAQAEQVILEQANKRRRAQILDCAETAEQERTKIVAQAQAEI 109
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 117 QALKEQLQAQMPVFEDELNKRVM 140
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 110 EAEKRRAR-----EELRKQVAI 126
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

RESULT 10
LWEBC
H+-transporting two-sector ATPase (EC 3.6.3.14) chain b - Escherichia coli (strain K-12)
N:Alternate names: hydrogen ion-transporting ATP synthase chain b; proton-translocating
C:Species: Escherichia coli

C;Date: 02-Apr-1982 #sequence_revision 06-Jul-1982 #text_change 01-Mar-2002
 C;Accession: D93732; B90104; B90101; I41273; I55242; A65177; T45004; A01070
 R;Gay, N.J.; Walker, J.E.
 Nucleic Acids Res. 9, 3919-3926, 1981
 A;Title: The *atp* operon: nucleotide sequence of the promoter and the genes for the membrane-transporting ATP synthase chain I
 A;Reference number: A93732; MUID:82059437; PMID:6272190
 A;Accession: D93732
 A;Molecule type: DNA
 A;Residues: 1-156 <GAY>
 A;Cross-references: GB:V00264; GB:X00771; NID:941023; PIDN:CAA23516.1; PID:g581045
 R;Kanazawa, H.; Mabuchi, K.; Kayano, T.; Nouri, T.; Sekiya, T.; Futai, M.
 Biochem. Biophys. Res. Commun. 102, 613-620, 1981
 A;Title: Nucleotide sequence of the genes for F-0 components of the proton-translocating ATP synthase
 A;Reference number: A90104; MUID:82134799; PMID:6277311
 A;Accession: B90104
 A;Molecule type: DNA
 A;Residues: 1-32, F', 34-71, D', 73-131 <KAN>
 R;Mabuchi, K.; Kanazawa, H.; Kayano, T.; Futai, M.
 Biochem. Biophys. Res. Commun. 102, 172-179, 1981
 A;Title: Nucleotide sequence of the gene coding for the delta subunit of proton-translocating ATP synthase
 A;Reference number: A90101; MUID:82068433; PMID:6458296
 A;Accession: B90101
 A;Molecule type: DNA
 A;Residues: 132-156 <MAB>
 A;Cross-references: GB:M12212; GB:M12213; NID:g145396; PIDN:AAA20043.1; PID:g145397
 R;Kanazawa, H.; Futai, M.
 Ann. N. Y. Acad. Sci. 402, 45-64, 1982
 A;Title: Structure and function of H⁺-ATPase: What we have learned from *Escherichia coli*
 A;Reference number: I41271; MUID:83176724; PMID:6301339
 A;Accession: I41271
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-156 <RES>
 A;Cross-references: GB:M25464; NID:g146318; PIDN:AAA83871.1; PID:g146321
 R;Porter, A.C.G.; Kumamoto, C.; Aldape, K.; Simoni, R.D.
 J. Biol. Chem. 260, 8182-8187, 1985
 A;Title: Role of the b subunit of the *Escherichia coli* proton-translocating ATPase: A mutation in the b subunit of the *Escherichia coli* proton-translocating ATPase
 A;Reference number: I55242; MUID:85234519; PMID:2861200
 A;Accession: I55242
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-61, N', 63-156 <RES>
 A;Cross-references: GB:M10422; NID:g148144; PIDN:AAA24741.1; PID:g148145
 R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A;Title: The complete genome sequence of *Escherichia coli* K-12.
 A;Reference number: A64720; MUID:97426617; PMID:9278503
 A;Accession: A65177
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-156 <BLAT>
 A;Cross-references: GB:AE000450; GB:U00096; NID:g1790166; PIDN:AACT6759.1; PID:g1790174;
 R;Nielsen, J.; Hansen, F.G.; Hoppe, J.; Friedl, P.; Von Meyenburg, K.
 Mol. Gen. Genet. 184, 33-39, 1981
 A;Title: The nucleotide sequence of the *atp* genes coding for the F-0 subunits a, b, c and sigma 3 of the membrane-transporting ATP synthase chain I
 A;Reference number: Z22893; MUID:82147764; PMID:6278247
 A;Accession: T45004
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-156 <NIE>
 A;Cross-references: EMBL:V00266; NID:g41031; PIDN:CAA23523.1; PID:g581046
 C;Genetics:
 A;Gene: *atpF*, *uncP*
 A;Map position: 84 min
 A;Start codon: GTG
 C;Complex: this is one of the three chains of the nonenzymatic membrane component (F₀) of the membrane-transporting ATP synthase chain I
 C;Superfamily: H⁺-transporting ATP synthase chain I
 C;Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; transmembrane protein

Query Match 17.3%; Score 123; DB 1; Length 156;
 Best Local Similarity 27.1%; Pred. No. 0.074;

Matches 39; Conservative 27; Mismatches 56; Indels 22; Gaps 4;
 QY 1 MNISVNPYLMAVVFVLLVMNVMVYRPLAFMNRQAEIKDSIAKIKTDNAQSVFI 60
 DQ 1 MNLNATILGQATAFVFLVFCMK---YVPPMLAAIEKQKEIADGLA-----SAER 49
 QY 61 GHOIEALLKEAAEKRRRIIAEIAIOKATESYDAVIKQENELNQFEAFKQLONEKQALK 120
 DQ 50 AHKDLDLAKASA-----TDLKKAKAEAOVIEQANKRSQILDEAKAEAEQERTKIV 102
 QY 121 EQLQAMQPV-----FEDELNKRVM 140
 DQ 103 AQAQAEIEAEKRAEERLEKQVAI 126
 RESULT 11
 F91213
 membrane-bound ATP synthase subunit b *atpF* [imported] - *Escherichia coli* (strain O157:H7)
 C;Species: *Escherichia coli*
 C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C;Accession: F91213
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gaasawa, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genome
 A;Reference number: A99629; MUID:21156231; PMID:11258796
 A;Accession: F91213
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-156 <HAY>
 A;Cross-references: GB:BA000007; PIDN:BA038101.1; PID:g13364153; GSPDB:GN00154
 A;Experimental source: strain O157:H7, substrain RMD 050952
 C;Genetics:
 A;Gene: *ECs4678*
 C;Superfamily: H⁺-transporting ATP synthase chain I
 Query Match 17.3%; Score 123; DB 2; Length 156;
 Best Local Similarity 27.1%; Pred. No. 0.074;
 Matches 39; Conservative 27; Mismatches 56; Indels 22; Gaps 4;
 QY 1 MNISVNPYLMAVVFVLLVMNVMVYRPLAFMNRQAEIKDSIAKIKTDNAQSVFI 60
 DQ 1 MNLNATILGQATAFVFLVFCMK---YVPPMLAAIEKQKEIADGLA-----SAER 49
 QY 61 GHOIEALLKEAAEKRRRIIAEIAIOKATESYDAVIKQENELNQFEAFKQLONEKQALK 120
 DQ 50 AHKDLDLAKASA-----TDLKKAKAEAOVIEQANKRSQILDEAKAEAEQERTKIV 102
 QY 121 EQLQAMQPV-----FEDELNKRVM 140
 DQ 103 AQAQAEIEAEKRAEERLEKQVAI 126
 RESULT 12
 G8059
 membrane-bound ATP synthase subunit b *atpF* [imported] - *Escherichia coli* (strain O157:H7)
 C;Species: *Escherichia coli*
 C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 27-Nov-2001
 C;Accession: G8059
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca
 Nature 409, 529-533, 2001
 A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A;Reference number: A85480; MUID:21074935; PMID:11206551
 A;Accession: G8059
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-156 <STO>
 A;Cross-references: GB:AE0005174; NID:g12518593; PIDN:AAG58939.1; GSPDB:GN00145; UWGP:Z5
 A;Experimental source: strain O157:H7, substrain EDL933
 C;Genetics:
 A;Gene: *atpF*
 C;Superfamily: H⁺-transporting ATP synthase chain I

```

Query Match          17.3%; Score 123; DB 2; Length 156;
Best Local Similarity 27.1%; Pred. No. 0.074;
Matches 39; Conservative 27; Mismatches 56; Indels 22; Gaps 4;

Qy      1 MNISVNYPLMAVTVFVFVLWAMNVYRLLAFPMNRQAIEKDSLAKIKTDAQSVEI 60
        |||: |:|||||
Db      1 MNLNATILGQAIAFLVFLFCMK---YWPPLMAAIEKRQEIAADGLA-----SAER 49

Qy      61 GHQT'EALLKEAAEKKREIIAEIAIOKATESYDAVIKQKENELNQEFAPAKOLONEKQALK 120
        |||: |:|||||
Db      50 AHKDLDLAKASA-----TDQLKKAKAAEQVIIIEQANKRRSQILDENKAESAEOERTKIV 102

Qy      121 EQLQAQMFPV-----FEDELNKRYAM 140
        |||: |:|||||
Db      103 AQAQAEIEAEERKKRAEBELRKQVAI 126

RESULT 13
F84119
ATP synthase subunit b atpF [imported] - Bacillus halodurans (strain C-125)
C/Species: Bacillus halodurans
C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C/Accession: F84119
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hirai, T.
Nucleic Acids Res 28, 4317-4331, 2000
A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A/Reference number: A83650; MUID:20512582; PMID:11058132
A/Accession: F84119
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-162 <STO>
A/Cross-references: GB:AP001519; GB:BA000004; MID:g10176109; PIDN:BAB07477.1; GSPDB:GNOC000001
A/Experimental source: strain C-125
C/Genetics:
A/Gene: atpF
C/Superfamily: H+-transporting ATP synthase chain I

```

```

Query Match          17.2%; Score 122.5; DB 2; Length 162;
Best Local Similarity 23.5%; Pred.No. 0.084;
Matches 31; Conservative 36; Mismatches 56; Indels 9; Gaps 2;

QY      15 VVFVLLILWANNVWVYRPLLAFMNROQAEIIKDSLAKTKTDNAQSVFCHQTEALLKKAEEK 74
      14 IAFCVLLWLLSKFALPKPLMGVMEKREQMINDIQADKORKAAQOYLEQORLAVAKAREE 73
      75 RREIIIAEIQAKTESYDAVIKQENELNQEFQAFKQLQNEKQ-----ALKEQQLQA----- 125
      74 AQEIVQKAKKLSEQQQOEIVEAARAEGERLKEAALAEIQREKEQAVASLFEQVASLSVLI 133
      126 QMPVFEDELNKR 137
      134 ATKVIEKELDEK 145

RESULT 14
GB82952
ATP synthase B chain PA5558 [imported] - Pseudomonas aeruginosa (strain PAO1)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: GB82952
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br-
      adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
      ; Lory, S.; Olson, M.V.
      Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho-
      gen
A:Reference number: AB2950, UID:20437337; PMID:10984043
A:Accession: GB82952
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-156 <STO>
A:Cross-references: GB:AE004967; GB:AE004091; NID:G9951884; PIDN:AAG08943.1; GSPDB:GN001
A:Experimental source: strain PAO1

```

```

C:Genetics:
A:Gene: acpP; PA5558
C:Superfamily: H+-transporting ATP synthase chain I

Query Match          16.9%; Score 120.5; DB 2; Length 156;
Best Local Similarity 28.1%; Pred. No. 0.11; 49; Indels 21; Gaps 5;
Matches 38; Conservative 27; Mismatches

QY 1 MNISVNPYLMAVTVFVLLWMNVVYRPLAFMDNRQAEIKDSIAKIKTDNAQSVEI 60
    |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 1 MNINATLIGCSVAFFIVFLFCMK---FVMPVPIAALQERQKTIADGL-DAANRAARDLEL 56
    |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

QY 61 GHQIEALLKEAAEKREIIIAEATQKATESYDAVTKKENELNQEFFAFKQLQNEKQALK 120
    |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 57 AHE-----KAQQLREAKQAQE-----IVEQAKRANQIVDEARDQARTGEKRLK 102
    |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

QY 121 EQLOQAMPVFEDELN 135
    |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 103 AQAQAEI---EQELN 114
    |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

RESULT 15
S01399
H+-transporting two-sector ATPase (EC 3.6.3.14) chain b precursor - thermophilic bacter
C:Species: thermophilic bacterium PS-3
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 14-Dec-2001
C:Accession: S01399
Biochim. Biophys. Acta 933, 141-155, 1998
Mol. Cell. Biochem. 194, 131-140, 1999
A:Title: Sequence and over-expression of subunits of adenosine triphosphate synthase in
A:Reference number: S01397; PMID:88163679; PMID:2894854
A:Accession: S01399
A:Molecule type: DNA
A:Residues: 1-163 <C>H>
A:Cross-references: EMBL:X07804; NID:G45808; PIDN:CAA30650.1; PID:G581473
A:Note: part of this sequence, including the amino end of the mature protein, was confi
C:Genetics:

```

```
A:Start codon: GTG
C:Superfamily: H+-transporting ATP synthase chain I
C:C.keywords: ATP biosynthesis; hydrolase; membrane-associated complex; transmembrane protein; F1F1-163/Product: H+-transporting ATP synthase chain b #status experimental <NAT>

Query Match      16.9%; Score 120.5; DB 1; Length 163;
Best Local Similarity 22.7%; Pred No. 0.12; Indels 9; Gaps 2;
Matches 30; Conservative 36; Mismatches 57;

15 VVFVLLWAMNVVRPIFLAFMDNRQAEIKDSLAKITDQAOSVEIGHQTALLKEAAEK 74
   :::|||: :::|||: :::|||: :::|||: :::|||: :::|||: :::|||: :
27 LMFIIALLRRKFAQPLMNINKOREHIAITSTRKNDRQEAELLEEQLRLMKQRQE 86
   :::|||: :::|||: :::|||: :::|||: :::|||: :::|||: :::|||: :
75 RREIIAEAIQKATESYDAVIKQENLQEFAPAFKOLQNEKO-----ALKEQLQA---- 125
   :::|||: :::|||: :::|||: :::|||: :::|||: :::|||: :::|||: :
87 AQALIENASLAEEQKQEIIVASARAEAEKVKEAKKEIEREKEQMAALREQVASLVLI 146
   :::|||: :::|||: :::|||: :::|||: :::|||: :::|||: :::|||: :

126 QMPVFEDLEINR 137
    |||||
147 ASKVIKELTEQ 158
```

Search completed: December 10, 2003, 18:38:00
Job time : 12.6126 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 18:29:11 / Search time 5.65445 Seconds
(without alignments)
1197.613 Million cell updates/sec

Title: US-10-080-113-4

Perfect score: 711

Sequence: 1 MNISVNPYLMAVFWVFVLL.....AQMEVFEDLNKRVAMGLGS 144

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	143	20.1	159	1 ATPX_SYNPI	Q05367 synechococ
2	133	18.7	222	1 ATPX_SPIOL	P31853 spinacia ol
3	123	17.3	156	1 ATPF_ECOLI	P00859 escherichia
4	122.5	17.2	162	1 ATPF_BACHD	Q36811 bacillus ha
5	120.5	16.9	163	1 ATPF_BACF3	P09221 bacillus ps
6	119	16.7	159	1 ATPF_THIPE	P41172 thiobacillu
7	119	16.7	168	1 ATPF_PROMO	P21904 propionigen
8	117	16.5	163	1 ATPX_GUITH	O78478 guillardia
9	116.5	16.4	159	1 ATPX_GUITH	Q02852 antithamnio
10	116.5	16.4	163	1 ATPX_OCHNE	Q04608 ochrosphaer
11	116	16.3	158	1 ATPX_PORPU	P31245 porphyra pu
12	113.5	16.0	164	1 ATPX_CYAPA	P48085 cyanophora
13	112	15.8	164	1 ATPF_RICCN	Q92J23 rickettsia
14	110.5	15.5	163	1 ATPF_BACFI	P22481 bacillus fi
15	109.5	15.4	157	1 ATPX_GALSU	P35012 galdieria s
16	108	15.2	3911	1 AKAG_HUMAN	Q99996 h a-kinase
17	107	15.0	156	1 ATPF_VIBAL	P22989 vibrio algi
18	107	15.0	163	1 ATPF_EUCAP	Q51876 buchnera ap
19	106.5	15.0	162	1 ATPF_BACCA	P41014 bacillus ca
20	106	14.9	156	1 ATPF_VIBCH	Q9knh1 vibrio chol
21	106	14.9	161	1 ATPF_EUCAI	P57120 buchnera ap
22	106	14.9	163	1 ATPX_ANASP	P2410 anabaena sp
23	105	14.8	167	1 ATPF_RICPR	Q32ec4 rickettsia
24	104.5	14.7	143	1 ATPF_SYN3	P27183 synechocyst
25	103.5	14.6	207	1 ATPF_MYCPN	Q50327 mycoplasma
26	103	14.5	158	1 ATPX_SYN6	P08446 synechococ
27	101	14.2	1690	1 C190_DROME	Q9vj65 drosophila
28	101	14.2	1941	1 MYH2_HUMAN	Q9ukx2 homo sapien
29	99.5	14.0	1875	1 MLP1_YEAST	Q02455 saccharomyc
30	99	13.9	170	1 ATPF_BACSU	P37814 bacillus su
31	98.5	13.9	1818	1 HMW2_MYCPN	P75471 mycoplasma
32	98	13.8	156	1 ATPF_HABIN	P43720 haemophilus
33	98	13.8	627	1 DTNB_HUMAN	O60941 homo sapien

34 97.5 13.7 757 1 IMMT_MOUSE
35 97.5 13.7 845 1 SCPI_MESAU
36 97 13.6 810 1 NFM_BOVIN
37 97 13.6 845 1 NFM_RAT
38 97 13.6 915 1 NFM_HUMAN
39 96.5 13.6 172 1 ATPF_BACME
40 95.5 13.4 790 1 MUS2_THETN
41 95 13.4 454 1 NFM_FIG
42 95 13.4 644 1 NFM_RABIT
43 95 13.4 848 1 NFM_MOUSE
44 95 13.4 1140 1 RA18_SCHPO
45 94.5 13.3 390 1 EMRA_ECOLI

ALIGNMENTS

RESULT 1
ATPX_SYNPI
ID ATPX_SYNPI STANDARD; PRT; 159 AA.
AC Q05367,
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP synthase B' chain (EC 3.6.3.14) (Subunit II).
GN ATPG.
OS Synechococcus sp. (strain PCC 6716).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32048;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93371369; PubMed=8363578;
RA van Walraven H.S., Lutter R., Walker J.E.;
RT "Organization and sequences of genes for the subunits of ATP synthase
in the thermophilic cyanobacterium Synechococcus 6716.";
RL Biochem. J. 294:239-251(1993).
CC -!- FUNCTION: THE B'-SUBUNIT IS A DIVERGED AND DUPLICATED FORM OF
CC B FOUND IN PLANTS AND PHOTOSYNTHETIC BACTERIA.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SUBUNIT: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
CC (CF1O) SUBUNIT) OF THE ATPASE COMPLEX.
CC -!- SIMILARITY: BELONGS TO THE ATPASE B CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X70431; CAA49872.1; -
CC InterPro; IPR002146; ATPsyn B/B' sub.
CC Pfam; P500430; ATP-synt B; 1
CC Hydrogen ion transport; Transmembrane; CF(0).
CC SEQUENCE 159 AA; 17919 MW; EEAEE969BE28E4CE CRC64;
CC

Query Match
Best Local Similarity 27.7%; Score 143; DB 1; Length 159;
Matches 36; Conservative 34; Mismatches 46; Indels 14; Gaps 2;
QY 9 LNAVVFVVFVLLVWVWVYRPLAFMDNRQAEIKDLSAKIKTDNAQSVETGHQIE--- 65
31 LNAVVFVVFVLLVWVWVYRPLAFMDNRQAEIKDLSAKIKTDNAQSVETGHQIE 87
QY 66 -----ALLKEAEKREIIAEAIQKATESYDAVIQKKNELNQEFAPAKQIQNEKQ 117
88 AYTRREAQAIIEARAEAKQTATAETAAQQAQLAEMLKMAQAEIQQKQATQALEGQVS 147
QY 118 ALKEQLQAOQ 127
148 SLSEQLLAKL 157

```

RESULT 2
ATPX SPIOL STANDARD; PRT; 222 AA.
ID ATPX SPIOL STANDARD; PRT; 222 AA.
AC P31853; P32981;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP synthase B' chain, chloroplast precursor (EC 3.6.3.14) (Subunit
DE II).
GN ATPG.
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
OX NCBI_TaxID=3582;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=CV. Monatoli; TISSUE=Seedling;
RX MEDLINE=93314783; PubMed=8325369;
RA Herrmann R.G., Steppuhn J., Herrmann G.S., Nelson N.;
RT "The nuclear-encoded polypeptide cfo-II from spinach is a real, ninth
subunit of chloroplast ATP synthase."
RL FEBS Lett. 326:192-198(1993).
RN [2]
SEQUENCE OF 76-107.
RA Berzborn R.J., Otto J., Finke W., Meyer H.E., Block J.;
RT "Conclusions from N-terminal amino acid sequences of subunits delta
from spinach and maize CF-1 and of subunits I and II from spinach
CF-0."
RL Biol. Chem. Hoppe-Seyler 368:551-552(1987).
CC -1- FUNCTION: THE B'-SUBUNIT IS A DIVERGED AND DUPLICATED FORM OF
B FOUND IN PLANTS AND PHOTOSYNTHETIC BACTERIA.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
H(+) (Out).
CC -1- SUBUNIT: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
(CF(0) SUBUNIT) OF THE ATPASE COMPLEX.
CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC -1- SIMILARITY: BELONGS TO THE ATPASE B CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; X71397; CAA50520.1; -
CC PIR; S34473; S34473.
CC InterPro; IPR002146; ATPsynth_B' sub.
CC Pfam; PF00430; ATP-synt_B; 1.
CC KEGG; Hydrogen ion transport; Transmembrane; CF(0); Chloroplast;
CC Transit peptide.
CC TRANSIT 1 75 CHLOROPLAST.
CC CHAIN 76 222 ATP SYNTHASE B' CHAIN.
CC TRANSMEM 83 103 POTENTIAL.
CC CONFLICT 79 79 E > L (IN REF. 2).
CC SEQUENCE 222 AA; 24459 MW; 8664CAEA910A5ED1 CRC64;
Query Match 18.7%; Score 133; DB 1; Length 222;
Best Local Similarity 25.0%; Pred. No. 0.017;
Matches 33; Conservative 37; Mismatches 56; Indels 6; Gaps 3;
QY 10 MAUVFVUULLWAMVWYVRLPAFMDNRQAEIKDLSAKIKTDNAQSVGEIGHOIEALLK 69
DB 90 LPIMAEFLFPMALDKIYTPJGDFMDKRDASIKELQSLGVKDTSEVKQLSEQANAVNR 149
QY 70 EAAEKREITA--EATQKATE-SYDAVVKQKNELNQCFEFAFQKQNEKQALKEQLOAQ 126
DB 150 AA---RAEISAALNKKKKTGLEVEAKLAEGRRKIEVEALQALGSLQEQKEDTIKSLDSQ 206

```

```

QY 127 MPVFEDELKRV 138
DB 207 ISALSDDIVKVV 218

RESULT 3
ATPF ECOLI STANDARD; PRT; 156 AA.
ID ATPF ECOLI STANDARD; PRT; 156 AA.
AC P00859;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE ATP synthase B chain (EC 3.6.3.14)
GN ATPF OR UNCF OR PAPF OR B3736 OR C4664 OR Z5234 OR ECS4678 OR SF3816.
OS Escherichia coli O6,
OS Escherichia coli O157:H7, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992, 83334, 623;
RN [1]
SEQUENCE FROM N.A.
RC SPECIES=E.coli;
RX MEDLINE=85121806; PubMed=6395859;
RA Walker J.E., Gay N.J., Saraste M., Eberle A.N.;
RT "DNA sequence around the Escherichia coli unc operon. Completion of
the sequence of a 17 kilobase segment containing asnA, oriC, unc,
gms and phoS."
RL Biochem. J. 224:799-815(1984).
RN [2]
SEQUENCE FROM N.A.
RC SPECIES=E.coli;
RX MEDLINE=82059437; PubMed=6272190;
RA Gay N.J., Walker J.E.;
RT "The unc operon: nucleotide sequence of the promoter and the genes
for the membrane proteins, and the delta subunit of Escherichia coli
ATP-synthase."
RL Nucleic Acids Res. 9:3919-3926(1981).
RN [3]
SEQUENCE FROM N.A.
RC SPECIES=E.coli;
RX MEDLINE=82134799; PubMed=6277311;
RA Kanazawa H., Mabuchi K., Kayano T., Nomi T., Sekiya T., Futai M.;
RT "Nucleotide sequence of the genes for F0 components of the proton-
translocating ATPase from Escherichia coli: prediction of the primary
structure of F0 subunits."
RL Biochem. Biophys. Res. Commun. 103:613-620(1981).
RN [4]
SEQUENCE FROM N.A.
RC SPECIES=E.coli;
RX MEDLINE=82147764; PubMed=6278247;
RA Nielsen J., Hansen F.G., Hoppe J., Friedl P., von Meyenburg K.;
RT "The nucleotide sequence of the atp genes coding for the F0 subunits
a, b, c and the F1 subunit delta of the membrane bound ATP synthase
of Escherichia coli."
RL Mol. Gen. Genet. 184:33-39(1981).
RN [5]
SEQUENCE FROM N.A.
RC SPECIES=E.coli;
RX MEDLINE=85234519; PubMed=2861200;
RA Porter A.C.G., Kumamoto C., Aldape K., Simoni R.D.;
RT "Role of the b subunit of the Escherichia coli proton-translocating
ATPase. A mutagenic analysis."
RL J. Biol. Chem. 260:8182-8187(1985).
RN [6]
SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=KL2 / MG1655;
RX MEDLINE=93315143; PubMed=7686882;
RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
genome: organizational symmetry around the origin of replication."
RL Genomics 16:551-561(1993).

```

[7] SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.I.T., Donnenberg M.S., Blattner F.R.;
RT "extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
[8]
RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
[9]
RC SPECIES=FROM N.A.
RX MEDLINE=12156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kuhsara S., Shiba T., Hattori M., Shinagawa H.;
RT "complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
[10]
RC SPECIES=FROM N.A.
RX MEDLINE=132-156 FROM N.A.
RC SPECIES=E.coli;
RX MEDLINE=82068433; PubMed=6458296;
RA Nabuchi K., Kanazawa H., Kayano T., Futai M.;
RT "Nucleotide sequence of the gene coding for the delta subunit of
proton translocating ATPase of Escherichia coli.";
RL Biochem. Biophys. Res. Commun. 102:172-179(1981).
[11]
RC SPECIES=FROM N.A.
RX MEDLINE=S.flexneri; STRAIN=301 / Serotype 2a;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
[12]
RC MUTAGENESIS OF REGION GLU-77 THROUGH GLN-85.
RC SPECIES=E.coli;
RX MEDLINE=92041622; PubMed=1682301;
RA McCormick K.A., Cain B.D.;
RT "Targeted mutagenesis of the b subunit of F1F0 ATP synthase in
Escherichia coli: Glu-77 through Glu-85.";
RL J. Bacteriol. 173:7240-7248(1991).
[13]
RC STRUCTURE BY NMR OF 1-34.
RC SPECIES=E.coli;
RX MEDLINE=99269098; PubMed=10336456;
RA Dmitriev O., Jones P.C., Jiang W., Fillingame R.H.;
RT "Structure of the membrane domain of subunit b of the Escherichia
coli F0F1 ATP synthase.";
RL J. Biol. Chem. 274:15598-15604(1999).
CC -!- CATALYTIC ACTIVITY: ATP + H2O + H+(In) = ADP + phosphate +
H+(Out).
CC -!- SUBUNIT: F-TYPE ATPASS HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC


```

RT mutants."
RL FEMS Microbiol. Lett. 122:19-26 (1994).
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SIMILARITY: BELONGS TO THE ATPASE B CHAIN FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M81087; AAA53123.1; -
DR InterPro; IPR005864; ATPaseB.
DR InterPro; IPR002146; ATPsynt_B/B'sub.
DR Pfam; PF00430; ATP-synt_B; 1.
DR TIGRFAMS; TIGR01144; ATP_synt_b; 1.
KW Hydrogen ion transport; Transmembrane; CF(0).
FT TRANSMEM 4 POTENTIAL.
SQ SEQUENCE 159 AA; 17884 MW; C5CFE547398831A CRC64;

Query Match 16.7%; Score 119; DB 1; Length 159;
Best Local Similarity 27.3%; Pred. No. 0.1;
Matches 35; Conservative 29; Mismatches 29; Indels 2; Gaps 2;

QY 1 MN-1SVNPLYMAVVFVLLWMNVVYRLLAFMDNRQAEIKDSLAKIKTDNAQSVE 59
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1 MNPVGINGTLLI-VQLTFVILVALLYKMYGPLEKVMDDRAKIAADGLAAERKKEVAL 59
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 60 IGHOIEALLKEAAEKREIIAIAIQATESYDAVIKQENELNQEFAPAKQJQKQAL 119
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 60 AQRAVELVREAKDAKAEIINARERGVLRERAAQKAREADRIIAGARAEIDVETNRA 119
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 120 KEQLQAM 127
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 120 REVLRGQV 127
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 7

```

ATPF_PROMO STANDARD; PRT; 168 AA.
AC P21904;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP synthase B chain, sodium ion specific (EC 3.6.3.15).
GN ATPF OR UNCF.
OS Propionigenium modestum.
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Propionigenium.
CX NCBI_TaxID=2333;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 2376;
RX MEDLINE=9106741; PubMed=2174545;
RA Kaim G., Ludwig W., Dimroth P., Schleifer K.H.;
RT "Sequence of subunits a and b of the sodium ion translocating
RL adenosine triphosphate synthase of Propionigenium modestum.";
RL Nucleic Acids Res. 18:6697-6697 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 2376;
RX MEDLINE=92339434; PubMed=1386022;
RA Kaim G., Ludwig W., Dimroth P., Schleifer K.H.;
RT "Cloning, sequencing and in vivo expression of genes encoding the F0
RT part of the sodium-ion-dependent ATP synthase of Propionigenium
RT modestum in Escherichia coli.";

```

```

RL Eur. J. Biochem. 207:463-470 (1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 2376;
RX MEDLINE=91016937; PubMed=2170948;
RA Esser U., Krumholz L.R., Simoni R.D.;
RT "Nucleotide sequence of the F0 subunits of the sodium dependent F1F0
RN ATPase of Propionigenium modestum.";
RL Nucleic Acids Res. 18:5887-5888 (1990).
RN [4]
RP SEQUENCE OF 1-7.
RX MEDLINE=93138123; PubMed=8422943;
RA Garike U., Dimroth P.;
RT "N-terminal amino acid sequences of the subunits of the Na(+)-
RN translocating F1F0 ATPase from Propionigenium modestum.";
RL FEBS Lett. 316:189-92 (1993).
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- MISCELLANEOUS: THE ATPASE OF P. MODESTUM IS OF SPECIAL INTEREST
CC BECAUSE IT USES SODIUM IONS INSTEAD OF PROTONS AS THE
CC PHYSIOLOGICAL COUPLING ION.
CC -1- SIMILARITY: BELONGS TO THE ATPASE B CHAIN FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X54810; CAA38580.1; ALT_INIT.
DR EMBL; X66102; CAA46896.1; ALT_INIT.
DR EMBL; X33980; CAA37913.1; -
DR EMBL; X36461; CAA41370.1; ALT_INIT.
DR PIR; S23323; S23323.
DR InterPro; IPR005864; ATPases.
DR InterPro; IPR002146; ATPsynt_B/B'sub.
DR Pfam; PF00430; ATP-synt_B; 1.
DR TIGRFAMS; TIGR01144; ATP_synt_b; 1.
KW Sodium transport; Transmembrane; CF(0).
FT TRANSMEM 9 POTENTIAL.
SQ SEQUENCE 168 AA; 19201 MW; 461EA7572AF9ABDA CRC64;

Query Match 16.7%; Score 119; DB 1; Length 168;
Best Local Similarity 19.6%; Pred. No. 0.11;
Matches 30; Conservative 44; Mismatches 63; Indels 16; Gaps 3;

QY 1 MN1SVNPLYMAVVFVLLWMNVVYRPLAFMDNRQAEIKDSLAKIKTDNAQSVEI 60
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 9 VSDIDNFWQ---IINFILMEFFKKYFKPIAKVLDARKEKIANDLKQAEIDKEMAAKA 65
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 61 GHQIEALLKEAAEKREIIAIAIQATESYDAVIKQENELNQEFAPAKQJQKQAL 120
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 66 NGEAQGIVKSAKTEANEMLRRAEKXADERKETILK---EANTQREKMLKSAEVEIERQK 121
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 121 EQLQAM 144
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 122 EQARKELQELVDLAVKVAEKMKINEKVDKAKIGA 154
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
ATPX_GUITH STANDARD; PRT; 163 AA.
AC O78478;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP synthase B' chain (EC 3.6.3.14) (subunit II).
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 2376;
RX MEDLINE=92339434; PubMed=1386022;
RA Kaim G., Ludwig W., Dimroth P., Schleifer K.H.;
RT "Cloning, sequencing and in vivo expression of genes encoding the F0
RT part of the sodium-ion-dependent ATP synthase of Propionigenium
RT modestum in Escherichia coli.";

```

OS Guillardia theta (Cryptomonas phi).
 OG Chloroplast.
 OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
 OX NCBI_TaxID=55529;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lettsch C.E.W., Kowalik K.V., Douglas S.E.;
 RT "The atpa gene cluster of a cryptomonad, Guillardia theta: a piece in
 RT the puzzle of chloroplast genome development.";
 RL J. Phycol. 35:128-135(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99128221; PubMed=9929392;
 RA Douglas S.E., Penny S.L.;
 RT "The plastid genome of the cryptophyte alga, Guillardia theta:
 RT complete sequence and conserved synteny groups confirm its common
 RT ancestry with red algae.";
 RL J. Mol. Evol. 48:236-244(1999).
 CC -!- FUNCTION: THE B'-SUBUNIT IS A DIVERGED AND DUPLICATED FORM OF
 CC B FOUND IN PLANTS AND PHOTOSYNTHETIC BACTERIA.
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
 CC H(+)(Out).
 CC -!- SUBUNIT: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
 CC (CF(0) SUBUNIT) OF THE ATPASE COMPLEX.
 CC -!- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
 CC -!- SIMILARITY: BELONGS TO THE ATPASE B CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF041468; AAC35669.1; -
 CC InterPro: IPR002146; ATPsynt_B/B'sub.
 CC Pfam: PF00430; ATP-synt B; 1.
 CC Hydrogen ion transport; Transmembrane; CF(0); Chloroplast.
 CC SEQUENCE 163 AA; 18329 MW; D3D84A154FD5A94C CRC64;
 CC -----
 CC Query Match 16.5%; Score 117; DB 1; Length 163;
 CC Best Local Similarity 24.1%; Pred. No. 0.14;
 CC Matches 32; Conservative 36; Mismatches 45; Indels 20; Gaps 3;
 CC -----
 CC 9 LMAVVFVVFVLLWAMVWVYRPLAFMDNRQAEIKDSIAKIKTDNAQSVIGHQIEALL 68
 CC 30 LMAVQILLFWIL---NAVFNPVAKVLDREYVIRKNTQASDILAKAEATKQYE--- 83
 CC -----
 CC 69 KEAEKRE-----IIAAIQKATESYDAVTKOKENELNQEFAFAKQLOK 114
 CC 84 KDLAQEREAQIIISVAQEAQDIVALEIKQAQKDTLLVNEATSQLNSQKQKALSALD 143
 CC -----
 CC 115 EQQALKEQLOAQM 127
 CC 144 QVNTLTQIKSKL 156
 CC -----
 CC RESULT 9
 CC ATPX ANTSP STANDARD; PRT; 159 AA.
 CC ID ATPX ANTSP
 CC AC Q02852;
 CC DT 01-JUL-1993 (Rel. 26, Created)
 CC DT 01-JUL-1993 (Rel. 26, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE ATP synthase B' chain (EC 3.6.3.14) (Subunit II).
 CC GN ATPG.
 CC OS Antithamion sp.
 CC OG Chloroplast.
 CC OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Ceramiaceae;
 CC OX Antithamion.
 CC OX NCBI_TaxID=2767;
 CC RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=LB 95.79;
 RX MEDLINE=93021132; PubMed=1404401;
 RA Kostrewa M., Zetsche K.;
 RT "Large ATP synthase operon of the red alga Antithamion sp. resembles
 RT the corresponding operon in cyanobacteria.";
 RL J. Mol. Biol. 227:961-970(1992).
 CC -!- FUNCTION: THE B'-SUBUNIT IS A DIVERGED AND DUPLICATED FORM OF
 CC B FOUND IN PLANTS AND PHOTOSYNTHETIC BACTERIA.
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
 CC H(+)(Out).
 CC -!- SUBUNIT: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
 CC (CF(0) SUBUNIT) OF THE ATPASE COMPLEX.
 CC -!- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
 CC -!- SIMILARITY: BELONGS TO THE ATPASE B CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X63382; CAA44981.1; -
 CC Pfam: S26959; S26959.
 CC InterPro: IPR002146; ATPsynt_B/B'sub.
 CC Pfam: PF00430; ATP-synt B; 1.
 CC Hydrogen ion transport; Transmembrane; CF(0); Chloroplast.
 CC SEQUENCE 159 AA; 17833 MW; D48525443280B07 CRC64;
 CC -----
 CC Query Match 16.4%; Score 116.5; DB 1; Length 159;
 CC Best Local Similarity 24.6%; Pred. No. 0.15;
 CC Matches 31; Conservative 32; Mismatches 3; Indels 3; Gaps 1;
 CC -----
 CC 9 LMAVVFVVFVLLWAMVWVYRPLAFMDNRQAEIKDSIAKIKTDNAQSVIGHQIEALL 68
 CC 30 LMAQFLALTITL---NLVYKPLGKILDEREYNTAASALSKANDLTRKYEODL 86
 CC -----
 CC 69 KEAEKREIIIAEQKATESYDAVTKOKENELNQEFAFAKQLOKNEQALKEQLOAQM 128
 CC 87 AERKKAQDILIKVAQDAQNIYSSKIKERAKQADQMLNMENTYDQNLNKEQALNLEKQVD 146
 CC 129 VFEDL 134
 CC 147 ILSNQI 152
 CC -----
 CC RESULT 10
 CC ATPX OCHNE
 CC ID ATPX OCHNE STANDARD; PRT; 163 AA.
 CC AC Q40608;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE ATP synthase B' chain (EC 3.6.3.14) (Subunit II).
 CC GN ATPG.
 CC OS Ochrosphaera neapolitana.
 CC OG Chloroplast.
 CC OC Eukaryota; Haptophyceae; Isochrysidales; Ochrosphaera.
 CC OX NCBI_TaxID=351137;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=CCMP 593;
 CC RA Huss V.A.R., Tietze A.C., Julius C.;
 CC Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: THE B'-SUBUNIT IS A DIVERGED AND DUPLICATED FORM OF
 CC B FOUND IN PLANTS AND PHOTOSYNTHETIC BACTERIA.
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
 CC H(+)(Out).
 CC -!- SUBUNIT: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
 CC (CF(0) SUBUNIT) OF THE ATPASE COMPLEX.
 CC -!- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.

```

CC -!- SIMILARITY: BELONGS TO THE ATPASE B CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X99078; CAA67537.1;
CC DR InterPro: IPR002146; ATPsynt_B/B'sub.
CC DR Pfam: PF00430; ATP-synt_B; 1.
CC KW Hydrogen ion transport; Transmembrane; CF(0); Chloroplast.
CC SQ SEQUENCE 163 AA; 17987 MW; ADD99B80FA7015B1 CRC64;

Query Match
Best Local Similarity 26.2%; Score 116.5; DB 1; Length 163;
Matches 34; Conservative 32; Mismatches 57; Indels 7; Gaps 2;

QY 9 LMAVVFVLLWVNVVYRPLAFMDNRQAEIKDSLAKIKTDNAQSVIEIGHQIEAL 68
DB 30 LMAQFILLVTLTVF---FYKPIGNLLEEREAYINGNLSASAKLLQADELCQYEQ 86
QY 69 KEAEKRRREILAEATOKATESYDAVIKQENELNOEFPAFAKOLNEKQALKEOLOAOMP 128
DB 87 KQAKADAGSCADATEAKQVVALEAQAARKDASLVQVKNLEAQLKALQLEAQI- 145
QY 129 VFEDELNKRVR 138
DB 146 ---DELSQLI 152

RESULT 11
ATPX_PORPU STANDARD; PRT; 158 AA.
AC P51245;
DT 01-OCT-1996 (Rel. 34; Created)
DT 01-OCT-1996 (Rel. 34; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE ATP synthase B' chain (EC 3.6.3.14) (subunit II).
GN ATPG.
OS Porphyra purpurea.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Porphyra.
OX NCBI_TaxID=2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Avonport;
RA Reith M.E., Munnholland J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RL genome."
RL Plant Mol. Biol. Rep. 13:333-335 (1995).
CC -!- FUNCTION: THE B'-SUBUNIT IS A DIVERGED AND DUPLICATED FORM OF
CC B FOUND IN PLANTS AND PHOTOSYNTHETIC BACTERIA.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SUBUNIT: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
CC (CF(0) SUBUNIT) OF THE ATPASE COMPLEX.
CC -!- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC -!- SIMILARITY: BELONGS TO THE ATPASE B CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U38804; AAC08131.1;
CC DR InterPro: IPR002146; ATPsynt_B/B'sub.

```

```

DR Pfam: PF00430; ATP-synt B; 1.
KW Hydrogen ion transport; Transmembrane; CF(0); Chloroplast.
SQ SEQUENCE 158 AA; 17768 MW; FEC1F6FF6086DB3D CRC84;

Query Match
Best Local Similarity 25.4%; Score 116; DB 1; Length 158;
Matches 33; Conservative 34; Mismatches 49; Indels 14; Gaps 3;

QY 9 LMAVVFVLLWVNVVYRPLAFMDNRQAEIKDSLAKIKTDNAQSVIEIGHQIE--- 65
DB 25 LMAQFILLVTLTVF---NTIFYKPVTKILDERDYIITLTASSMLVKADLAAYEEDL 81
QY 66 -----ALLKEAEKRR--REIIAEAIQKATESYDAVIKQENELNOEFPAFAKOLNEKQ 117
DB 82 SKARRDAQAIAASQDAQSVISDEIKKAQMAEKLITEASKQLNIQKEALKTLEDQVD 141
QY 118 ALKEOLOAQM 127
DB 142 TLDQIKTKL 151

RESULT 12
ATPX_CVAPA STANDARD; PRT; 164 AA.
AC P48085;
DT 01-FEB-1996 (Rel. 33; Created)
DT 01-FEB-1996 (Rel. 33; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE ATP synthase B' chain (EC 3.6.3.14) (subunit II).
GN ATPG.
OS Cyanophora paradoxa.
OG Cyanelle.
OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX NCBI_TaxID=2762;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
RA Bryant D.A.;
RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa."
RL Plant Mol. Biol. Rep. 13:327-332 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA Farley J.Y., Schlucher W.M., Chung S., Newmann-Spallart C.,
RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RL the genetic complexity of a primitive plastid."
RL (in) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL Schwenmler W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL (1997).
CC -!- FUNCTION: THE B'-SUBUNIT IS A DIVERGED AND DUPLICATED FORM OF
CC B FOUND IN PLANTS AND PHOTOSYNTHETIC BACTERIA.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SUBUNIT: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
CC (CF(0) SUBUNIT) OF THE ATPASE COMPLEX.
CC -!- SUBCELLULAR LOCATION: Cyanelle thylakoid membrane.
CC -!- SIMILARITY: BELONGS TO THE ATPASE B CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U30821; AAA81256.1;
CC DR InterPro: IPR002146; ATPsynt_B/B'sub.

```

54 -----NAQSV EIGHQIEALLKEAAEKRRREIIAEAIQKATESYDAVIKQKENELNQEF 105

RESULT 15

```
ATPX_GALSU          STANDARD;          PRT;   157 AA.
ID  ACPX_GALSU
AC  P35012;
DT  01-FEB-1994 (Rel. 28, Created)
DT  01-FEB-1994 (Rel. 28, Last sequence update)
DT  15-SEP-2003 (Rel. 42, Last annotation update)
DE  ATP synthase B' chain (EC 3.6.3.14) (Subunit II).
GN  ATPG.
OS  Galdieria sulphuraria (Red alga).
OC  Chloroplast.
OC  Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
OC  Galdieria.
OX  NCBI_TaxID=130081;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=14-1-1 / Isolate 107.79/Goettingen;
RX  MEDLINE=94033298; PubMed=8219057;
RA  Kozlowska M., Zetsche K.;
RT  "Organization of plastid-encoded ATPase genes and flanking regions
RT  including homologues of infB and tsf in the thermophilic red alga
RT  Galdieria sulphuraria.";
RL  Plant Mol. Biol. 23:67-76(1993).
CC  -!- FUNCTION: THE B'-SUBUNIT IS A DIVERGED AND DUPLICATED FORM OF
CC  B FOUND IN PLANTS AND PHOTOSYNTHETIC BACTERIA.
CC  -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC  H(+) (Out).
CC  -!- SUBUNIT: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
CC  (CF(0) SUBUNIT) OF THE ATPASE COMPLEX.
CC  -!- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC  -!- SIMILARITY: BELONGS TO THE ATPASE B CHAIN FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X67814; CAA48022.1; -.
DR  InterPro; IPR002146; ATPsynt_B/B'sub.
DR  Pfam; PF00430; ATP-synt_B; 1.
KW  Hydrogen ion transport; Transmembrane; CF(0); Chloroplast.
SQ  SEQUENCE 157 AA; 18186 MW; 2DBE0F8C073FD748 CRC64;

Query Match          15.4%; Score 109.5; DB 1; Length 157;
Best Local Similarity 22.7%; Pred. No. 0.44;
Matches 30; Conservative 38; Mismatches 63; Indels 1; Caps 1;

Qy  10 MAVYVVFVLLMANNVYRPLAFMDNRQAEIKDSLAKIKTDNAQSVIGHQIEALK 69
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  24 LSFIALEFLILTSVNLIIYQPISKVIDSREDYIRENLKASLYLDQANLTKKYLELI 83
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy  70 EAAREKREITAEATQKATESDAVIKQENELNQEFAFAKQLCNEKQALKQEQMPV 129
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  84 TARKEAIKNVTSQTEACEFVNQALISQAQKEAQQLIOSSMMQFKEKKNKAIYSLKQVEQ 143
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy  130 FEDEL-NKRVAM 140
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  144 LSEQIKNKLISI 155
```

Search completed: December 10, 2003, 18:34:39
Job time : 7.65445 secs

